

**Classroom-based Assessment of Young EFL Learners in the
Chinese Context: Teachers' Conceptions and Practices**

Qiaozhen Yan

A thesis submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy (Ph.D.) in Education (Applied
Linguistics & TESOL)

The University of Auckland, Auckland
New Zealand

2020

ABSTRACT

This study investigated primary school English as a foreign language (EFL) teachers' conceptions and practices regarding classroom-based assessment (CBA) in China. It had four objectives: (1) to examine teachers' conceptions about the purposes and processes of classroom-based assessment; (2) to examine the effects of teacher attributes and work environments on teachers' conceptions; (3) to examine the current status of teachers' classroom-based assessment practices; and (4) to investigate the relationships between teachers' classroom-based assessment conceptions and practices.

An explanatory sequential mixed-methods design was adopted. In Phase One, a survey study was carried out with 195 teachers to gain a general understanding of teachers' conceptions and practices regarding classroom-based assessment. In the subsequent Phase Two, three participants formed a case study, in which teachers' actual classroom assessment practices were observed, and rationales for their practices were elicited through post-observation interviews. Additional data were generated through the collection of relevant documents to provide complementary information on teachers' assessment practices.

The findings indicate that the teachers expressed strong and positive perceptions about the learning and teaching improvement purpose of classroom-based assessment; they were positively disposed to the four components of classroom-based assessment: planning assessment, collecting learning evidence through multiple methods, making professional judgments and providing descriptive feedback. The findings also suggest that teaching experience and school type were two critical influences on teachers' CBA conceptions.

The study suggests that the teachers had not fully applied CBA principles to support young learners' learning in their practices. Their instruction was not aligned with the curriculum standards, and learning objectives and success criteria were not transparent to students. Despite that the teachers used spontaneous assessment opportunities (e.g., questioning and observation) to modify teaching and provide immediate feedback to students, they relied heavily on formal assessment tasks and used these tasks for summative purposes. Student-involving assessment opportunities (e.g., self-and peer assessment) were seldom employed and, if used at times, they were mainly for summative purposes. There was a reliance on norm-referenced assessment rather than criterion-referenced or pupil-referenced assessment. The teachers frequently provided evaluative feedback rather than descriptive feedback. Overall, the teachers' classroom-based assessment practices were teacher-dominated, with students playing a passive role.

The relationships between teachers' conceptions and practices were complex. While teachers' conceptions of classroom-based assessment played an influential role in shaping their assessment practices, there were clear discrepancies between their CBA conceptions and practices. The discrepancies appeared to be due to a variety of factors, including teacher-related factors (e.g., teachers' expertise in classroom-based assessment and the tension between their core and peripheral conceptions), student-related factors (e.g., students' learning needs and parents' expectations), and contextual factors (e.g., class size, curriculum demands, school's assessment policies, and examination-driven culture).

It is hoped that this study will make a valuable contribution to research on classroom-based assessment from the perspective of teacher cognition, and offer insights into classroom-based assessment of young language learners. A number of

practical implications are discussed in relation to policymaking and language teacher education.

DEDICATION

To my parents

ACKNOWLEDGMENTS

This thesis would never have been possible without the support and guidance of a group of people. I take this opportunity to acknowledge their support gratefully.

I owe an enormous debt of gratitude to my main supervisor, Professor Lawrence Zhang, for giving me the precious opportunity to conduct my Ph.D. project under his supervision and leading me to the field of Applied Linguistics and TESOL. I am thankful for his academic support, scientific discussion, and great patience in guiding me through this project. His enthusiasm for research is contagious and motivational for me, especially during tough times in the thesis journey. I would also like to extend my sincere gratitude to my co-supervisor, Associate Professor Helen Dixon. She has inspired me with her wealth of knowledge of Assessment for Learning. It has been my great fortune to have both of them as my supervisors. I appreciate all their contributions of ideas and time to make my Ph.D. experience productive.

I would like to express my sincere thanks to the primary school EFL teachers in China who participated in this project and who so willingly shared their thoughts and experiences.

A special thanks goes to Dr. Libby Limbrick for proofreading my thesis and Dr. Mohamed Alansari for offering statistical assistance. I am also grateful to the examiners, Dr. Yongcan Liu and Professor Chris Davison. Their valuable comments and feedback have helped to extend my thinking about my research. My gratitude also goes to Professor Xiangdong Gu, who introduced me to the field of Language Assessment nine years ago. I am thankful for her continuous support in the years that followed.

Further, I would like to express my sincere thanks to the China Scholarship Council and the University of Auckland for awarding me a joint doctoral scholarship, which made my doctoral study possible.

I would also like to extend my gratitude to a group of friends and fellow doctoral students for their continued support during the thesis journey. A special thanks goes to my friends, Vivi Li, Jiani Yun, Meng Deng, Jing Sun, Molly Tian, Cen Chen, Suwen Sun, Jacob and Zijian Wang. How could I have kept moving on during the highs and lows of the journey without your support? Your faith in my ability to complete my doctoral study have encouraged me to overcome various challenges. Our friendship will be a treasure in my life.

Finally, I would like to extend my deepest gratitude to my loving family. Thank you, Mum and Dad. You have always been a constant source of love and support during my life.

TABLE OF CONTENTS

ABSTRACT	i
DEDICATION	iv
ACKNOWLEDGMENTS	v
TABLE OF CONTENTS	vii
LIST OF TABLES	xiii
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS	xvi
CHAPTER ONE INTRODUCTION	1
1.1 Chapter overview	1
1.2 Research context	1
1.2.1 English language education in primary schools in China	1
1.2.2 English language assessment in China.....	7
1.3 Theoretical background.....	10
1.4 Research questions	12
1.5 Significance of the study	12
1.6 Outline of the thesis	14
CHAPTER TWO LITERATURE REVIEW	16
2.1 Chapter overview	16
2.2 Classroom-based assessment	16
2.2.1 Paradigm shift in the assessment landscape	16
2.2.2 Definitions of classroom-based assessment	19
2.2.3 Purposes of classroom-based assessment.....	21
2.2.4 Principles of classroom-based assessment for learning	27
2.2.5 A framework for the practice of CBA.....	36
2.2.6 Research on classroom-based assessment in L2 contexts	44
2.3 Teacher conceptions.....	54
2.3.1 Origin of teacher conceptions research	54

2.3.2	Definitions of teacher conceptions.....	55
2.3.3	Importance of teacher conceptions.....	56
2.3.4	Teacher conceptions and classroom practice	57
2.3.5	Sources of teacher conceptions	59
2.3.6	L2 teachers' conceptions of classroom-based assessment	62
2.4	Assessment of young language learners	65
2.4.1	Definition and characteristics of YLLs	65
2.4.2	Research on YLL assessment.....	66
2.5	Summary	68
CHAPTER THREE METHODOLOGY.....		70
3.1	Chapter overview	70
3.2	Research questions.....	70
3.3	Philosophical worldview	71
3.4	Mixed-methods research design.....	72
3.4.1	Phase One: Survey study.....	73
3.4.2	Phase Two: Case study	75
3.5	Phase One — Survey study.....	77
3.5.1	Participants.....	77
3.5.2	Development of questionnaire	80
3.5.3	Data collection	87
3.5.4	Data analysis	88
3.6	Phase Two — Case study.....	93
3.6.1	Participants.....	93
3.6.2	Instruments.....	95
3.6.3	Data collection	99
3.6.4	Data analysis	101
3.7	Ethical considerations	105
3.7.1	Voluntary participation	105
3.7.2	Informed consent.....	105

3.7.3 Participants' rights to withdraw	106
3.7.4 Anonymity and confidentiality	106
3.8 Chapter summary	106
CHAPTER FOUR RESULTS OF THE SURVEY STUDY	107
4.1 Chapter overview	107
4.2 Exploring underlying constructs: CBA purposes, processes and practices.....	107
4.2.1 Classroom-based assessment purposes	107
4.2.2 Classroom-based assessment processes	110
4.2.3 Classroom-based assessment practices	113
4.3 Teachers' conceptions of classroom-based assessment	117
4.3.1 Teachers' conceptions of CBA purposes	117
4.3.2 Teachers' conceptions of CBA processes	118
4.3.3 Summary	121
4.4 Effect of teacher attributes and school environments on teachers' conceptions of CBA	122
4.4.1 Effect on teachers' conceptions of CBA purposes.....	123
4.4.2 Effect on teachers' conceptions of CBA processes.....	126
4.4.3 Summary	131
4.5 Teachers' classroom-based assessment practices	132
4.6 Relationships between teachers' conceptions and practices regarding classroom-based assessment.....	135
4.6.1 Relationships between teachers' perceived CBA purposes and their practices	135
4.6.2 Relationships between teachers' perceived CBA processes and their practices	140
4.6.3 Summary	145
4.7 Chapter summary	146
CHAPTER FIVE RESULTS OF THE CASE STUDY	148
5.1 Chapter overview	148
5.2 Planning assessment.....	148
5.2.1 Instructional objectives not guided by curriculum standards	148
5.2.2 Learning objectives not clearly communicated to students	150

5.2.3 Attempts made to explain assessment process rather than to clarify success criteria	153
5.2.4 Selection of assessment methods influenced by both students' needs and school' assessment policy	155
5.3 Collecting evidence of students' learning	156
5.3.1 Using spontaneous assessment opportunities.....	157
5.3.2 Using formal assessment tasks.....	165
5.3.3 Using student-involving assessment opportunities	169
5.4 Making professional judgments	175
5.4.1 Lacking criterion-referenced assessment	175
5.4.2 Pupil-referenced assessment occasionally used	176
5.4.3 High frequency of norm-referenced assessment	177
5.4.4 Checking reliability of judgments	179
5.5 Providing appropriate feedback	181
5.5.1 Descriptive feedback occasionally used to identify attainment and specify improvement	181
5.5.2 Evaluative feedback frequently used to enhance students' motivation.....	186
5.8 Chapter summary	191
CHAPTER SIX GENERAL DISCUSSION	193
6.1 Chapter overview	193
6.2 Teachers' conceptions of classroom-based assessment	194
6.2.1 Teachers' conceptions of CBA purposes	194
6.2.2 Teachers' conceptions of CBA processes	196
6.3 Influence of teacher attributes and work environments on teachers' conceptions of classroom-based assessment	201
6.3.1 Effect of teaching experience.....	201
6.3.2 Effect of previous education in assessment	202
6.3.3 Effect of school type	203
6.3.4 Effect of grade level.....	205
6.4 Teachers' classroom-based assessment practices	205

6.4.1 Planning assessment.....	206
6.4.2 Collecting evidence of students' learning.....	208
6.4.3 Making professional judgments.....	211
6.4.4 Providing appropriate feedback.....	212
6.4.5 Teachers' and students' roles in classroom-based assessment.....	213
6.5 Relationships between teachers' conceptions and practices regarding classroom-based assessment.....	214
6.5.1 Teacher-related factors.....	217
6.5.2 Student-related factors.....	219
6.5.3 Contextual factors.....	221
6.6 Chapter summary.....	224
CHAPTER SEVEN CONCLUSION, IMPLICATIONS, LIMITATIONS AND RECOMMENDATIONS	225
7.1 Chapter overview.....	225
7.2 Summary of major findings.....	225
7.3 Implications.....	227
7.3.1 Theoretical implications.....	227
7.3.2 Methodological implications.....	229
7.3.3 Practical implications.....	230
7.5 Limitations of the study.....	234
7.6 Recommendations for further research.....	236
7.7 Conclusion.....	237
REFERENCES	238
APPENDICES	281
Appendix 1 Primary School English Teachers' Classroom-based Assessment: Conceptions and Practices Questionnaire (English version).....	281
Appendix 2 Primary School English Teachers' Classroom Assessment: Conceptions and Practices Questionnaire (Chinese version).....	287
Appendix 3 Interview Guide.....	292
Appendix 4 Example of Coding Observation Data (Amy's class).....	295

Appendix 5 Descriptive Statistics for Items of Purpose Scale.....	306
Appendix 6 Descriptive Statistics for Items of Process Scale	307
Appendix 7 Descriptive Statistics for Items of Practice Scale.....	309
Appendix 8 Extract of Amy’s Teaching Plan	311

LIST OF TABLES

Table 3.1 Research Design for Survey Study	74
Table 3.2 Research Design for Case Study	77
Table 3.3 Demographic Information on Participants in Survey study.....	79
Table 3.4 Questionnaire Structure	80
Table 3.5 Major Statistical Analyses	90
Table 3.6 Demographic Information on Three Case Teachers.....	94
Table 3.7 Background Information on Classroom Observations.....	100
Table 4.1 EFA Results of Purpose Scale.....	108
Table 4.2 EFA Results of Process Scale.....	111
Table 4.3 EFA Results of Practice Scale.....	114
Table 4.4 Descriptive Statistics of Teachers' Conceptions of CBA Purposes.....	117
Table 4.5 Descriptive Statistics of Teachers' Conceptions of CBA Processes	119
Table 4.6 Comparison of Teachers' Conceptions of Three Assessment Methods.....	120
Table 4.7 Teachers' Conceptions of CBA Purposes by Teaching Experience.....	123
Table 4.8 Teachers' Conceptions of CBA Purposes by Previous Education in Assessment.....	124
Table 4.9 Teachers' Conceptions of CBA Purposes by School Type.....	125
Table 4.10 Teachers' Conceptions of CBA Purposes by Grade Level.....	126
Table 4.11 Teachers' Conceptions of CBA Processes by Teaching Experience.....	127
Table 4.12 Teachers' Conceptions of CBA Processes by Previous Education in Assessment.....	129
Table 4.13 Teachers' Conceptions of CBA Processes by School Type.....	130
Table 4.14 Teachers' Conceptions of CBA Processes by Grade Level.....	131

Table 4.15 Descriptive Statistics of Teachers' CBA Practices.....	132
Table 4.16 Comparison of Teachers' Practices of Three Assessment Methods.....	133
Table 4.17 Correlations of Teachers' Perceived CBA Purposes and their Practices....	136
Table 4.18 Predictions of Teachers' Perceived CBA Purposes on their Practices.....	137
Table 4.19 Correlations of Teachers' Perceived CBA processes and their Practices....	141
Table 4.20 Predictions of Teachers' Perceived CBA Processes on CBA practices....	142

LIST OF FIGURES

Figure 3.1 Overview of Research Design.....	73
Figure 3.2 Example of Generating the Pattern of Planning Assessment.....	104

LIST OF ABBREVIATIONS

AfL	Assessment for learning
ALP	Advanced Language Proficiency
CBA	Classroom-based assessment
CET	College English Test
CLT	Communicative Language Teaching
CSE	China's Standards of English
EFA	Exploratory factor analysis
EFL	English as a foreign language
ESL	English as a second language
IELTS	International English Language Testing System
KET	Cambridge English: Key for Schools
KMO	Kaiser-Meyer-Olkin's
L1	First language
L2	Second or foreign language
MOE	Ministry of Education
NECS	The new English Curriculum Standards
PET	Cambridge English: Preliminary for Schools
TBLT	Task-based Language Teaching
TEM	Test for English Majors
TOEFL	Test of English as a Foreign Language
YLE	Cambridge Young Learners of English
YLL	Young language learners

CHAPTER ONE

INTRODUCTION

1.1 Chapter overview

This thesis reports a study that investigated English as a foreign language (EFL) teachers' conceptions and practices regarding classroom-based assessment (CBA) of young language learners (YLLs) in China. This chapter begins with a description of the research context: English language education in primary schools and English language assessment in China. The theoretical background is established to identify major research gaps, leading to the research questions and the significance of the study. Finally, an outline of the thesis is presented.

1.2 Research context

This section provides a brief introduction to the context of the research from two aspects: English language education in primary schools and English language assessment in China. The introduction to primary school English language education in China provides background knowledge of national standards, teaching methods, assessment, textbooks, teacher supply, and class size. The introduction to English language assessment in China addresses the examination-driven culture, as well as the assessment reform advocated by the Chinese government.

1.2.1 English language education in primary schools in China

Teaching English to young learners is a major global trend for a variety of social, economic, and technological development reasons (Vettorel, 2014). This is no exception for China, where the starting age for formal English language education has been

lowered in recent decades (Butler, 2015). Since the Reform and Opening Policy of China in the late 1970s, English increasingly has been seen as crucial to modernization because it facilitates the development of scientific knowledge and technology (Jin & Cortazzi, 2002; Kirkpatrick & Liddicoat, 2017). Consequently, English has become a compulsory subject for junior secondary schools, senior secondary schools, and colleges and universities. At the beginning of the 21st century, there was a significant change in the English language education policy in China, with the Chinese Ministry of Education (MOE) issuing a document entitled the *Guidelines for Promoting Teaching English Courses in Primary Schools* (Ministry of Education, 2001b). According to this document, English language education should be compulsory from Grade 3 (age 8) of primary school in cities and suburban areas by September 2001 and then in rural areas by September 2002. In some large cities such as Beijing and Shanghai, English language education begins from Grade 1 of primary school. It has been estimated that there were over 60 million primary school EFL learners in China in 2012 (Gil, 2016).

1.2.1.1 National Standards for English

The MOE has made efforts to develop the new English Curriculum Standards (NECS), which was first promulgated in the *English Curriculum Standards for 9-year Compulsory Education and Senior Secondary Schools (Trial version)* (Ministry of Education, 2001a). While the 2001 curriculum standards covered both 9-year compulsory education (primary schools and junior secondary schools) and senior secondary schools, the standards were not specific. In 2003, the *English Curriculum Standards for Senior Secondary schools (Trial version)* (Ministry of Education, 2003), providing more specific curriculum standards, was developed for senior secondary schools. A decade later, a particular curriculum for compulsory education, the *English*

Curriculum Standards for Compulsory Education (2011 version) (Ministry of Education, 2011), was introduced.

The main aim of the NECS is to facilitate students' comprehensive language competence rather than merely help them master knowledge and skills (Gu, 2012). It conceptualizes English language competence in terms of five major domains: language skills (speaking, listening, reading and reading), language knowledge (pronunciation, vocabulary, grammar, functions, and topics), affect and attitudes (motivation, confidence, cooperation, patriotism, international perspectives), learning strategies (cognitive, meta-cognitive and communicative strategies, resourcing) and cultural awareness (cultural knowledge, understanding of English-speaking cultures, cross-cultural communication). The NECS encompasses nine competence-based levels. In China, the primary school consists of six grades, and, according to the NECS, students in Grade 4 are required to attain Level 1, and those in Grade 6 are required to attain Level 2. Specific learning targets are provided for all nine levels in the language skills domain, whereas specific targets for language knowledge, affect and attitudes, learning strategies, and cultural awareness are only provided for Levels 2 and 5.

In 2018 the MOE launched *China's Standards of English* (CSE), a unified standard for English learners or users across all proficiency levels in China, which provides another powerful impetus for the refinement of primary English education. The CSE works as a common reference for English language teaching, learning, and assessment in China (Liu & Pan, 2019; Ren & Liu, 2016). It describes Chinese learners' achievement in the target language, at different educational phases, from three perspectives: Language competences, linguistic knowledge, and strategies. More specifically, the CSE descriptors are divided into nine levels: CSE 1 and CSE 2 correspond to Grade 3 and Grade 6 primary school students, respectively; CSE 3 and

CSE 4 correspond to junior secondary school, and senior secondary school graduates, respectively; CSE 5 correspond to non-English major sophomores; CSE 6 correspond to English major sophomores or non-English major undergraduates; CSE 7 correspond to English major undergraduates; CSE 8 correspond to English major postgraduates; CSE 9 correspond to professional English users (e.g., professional translators and interpreters).

1.2.1.2 Teaching methods

Prior to the introduction of the NECS, English language education in China was dominated by the traditional grammar-translation teaching method (Zheng & Borg, 2014). This method emphasizes grammatical rules, words, and structures, paying little attention to students' communicative competence (Liao, 2004). The implementation of the NECS, which introduced Communicative Language Teaching (CLT) and Task-based Language Teaching (TBLT) to English teachers, represented an innovation in English language teaching methods in China (Bao, Zhang, & Dixon, 2016; Butler, 2011; Liu, Lin, & Wiley, 2016). As CLT aims to develop students' communicative competence in social interactions (Scrivener, 2011), students should be given opportunities to use language appropriately in a variety of real-life situations. In TBLT, students are encouraged to experience and learn the language by completing language tasks (Butler, 2011).

1.2.1.3 Assessment

With the introduction of the NECS, assessment became an important component of English language education in primary schools. The NECS proposed eight guidelines for assessment, seven of which focused on the use of formative assessment, thus

encouraging the role of formative classroom-based assessment in improving learning. The following are the eight guidelines for assessment, which are included in the NECS:

- (1) *Focus on the use of formative assessment for the purpose of student improvement*
- (2) *Reflect students' role in the process of assessment*
- (3) *Establish objectives for assessment and success criteria according to the NECS*
- (4) *Stress the appropriateness and variety of assessment methods*
- (5) *Formative assessment should be used to monitor and promote teaching and learning*
- (6) *Summative assessment should focus on the evaluation of students' comprehensive language use ability*
- (7) *Pay attention to the relationship between instruction and assessment*
- (8) *The central aim of assessment is to motivate students' interest* (Ministry of Education, 2011, pp. 33–38)

1.2.1.4 Textbooks

Textbooks are acknowledged as crucial to English language education in China (Ren & Han, 2016). They are not only a teaching instrument, but also reflect theories and methodology in language learning (Aliakbari & Jamalvandi, 2012; Ping, 2015). Therefore, developing good English textbooks has been an important issue for policy makers, and, for a long time, English textbooks were published only by the People's Education Press. As all provinces of China used a unified set of textbooks regardless of local differences in economy, English education, and culture, problems in teaching and learning arose. Currently, local governments are allowed to develop their own textbooks

in order to cater to the needs of different departments and schools (Wang, 2011). At the primary school level, for instance, there are around 30 sets of English textbooks. Some of the most popular textbooks include *PEP Primary English*, published by the People's Education Press, and *Standard English*, published by the Foreign Language Teaching and Research Press.

1.2.1.5 Teacher supply

The promotion of primary English education has also led to an increase in the need for EFL teachers in China. Although in 2012, there were over 300,000 primary EFL teachers (Gil, 2016), there is still a shortage of teachers due to a large number of primary EFL learners. According to Wen (2012), the teacher-student ratio at the primary school level, of about 1:160, is inadequate.

Moreover, not all primary EFL teachers are qualified. Teachers with limited knowledge of English language teaching and learning are often recruited. According to Wang (2009), fresh graduates with a degree in English from colleges and universities account for only a small proportion of teachers of primary English. The majority of primary EFL teachers are fresh graduates with a major that is not English or primary teachers who used to teach other subjects, such as Chinese and Mathematics, and, as such, lack relevant language knowledge. Furthermore, as these primary English teachers have not received formal training for teaching English to young learners, they often lack adequate pedagogical skills.

1.2.1.6 Class size

Large class size is another feature of primary school English language education. As suggested by Wang (2009), classes in Chinese primary schools are very large, with an average of 50 students. A large number of students in classes is considered a major

impediment to English language education in China; it constrains the use of communicative approaches, hinders teachers from taking account of students' needs, and makes it difficult for schools to organize in-service teacher training programs (Gil, 2016; Jin & Cortazzi, 1998; Zhu, 2003).

1.2.2 English language assessment in China

It is widely acknowledged that China has a long history of testing and examinations, which can be traced back to the imperial period, almost 2,000 years ago (Spolsky, 1995). The imperial examinations, used to select the highest officials for the nation in the Han Dynasty, are recognized by many scholars as the first standardized tests in the world (Bowman, 1989). The tradition of using examinations for selection has had a profound impact on the current education system (Yu & Jin, 2014), as students need to take a variety of standardized tests from primary to higher education. Examinations remain powerful influences in the social systems of China, including government official selection, professional promotion, and talent employment (Cheng, 2008; Cheng & Qi, 2006), which suggests that “being successful in examinations is the key to success in study, work and life” (Cheng & Curtis, 2010, p. 267).

Although the history of English language testing and examinations in China started late, around the 1860s, large-scale English examinations are evident at all levels of education (Adamson, 2004). English is one of the core subjects in secondary school and university entrance examinations. At the tertiary level, the national *College English Test* (CET) is taken by approximately 18 million students every year (G. Yu & Jin, 2016). The *Test for English Majors* (TEM), a criterion-referenced English proficiency test, is taken by English majors to determine whether they have met the required levels of

English language competence specified in the curriculum. English is also a compulsory subject for all undergraduate students who want to pursue a graduate degree in China.

A phenomenal number of Chinese learners take international English language tests. For example, the *Cambridge Young Learners of English* (YLE), introduced into China early in 1997, has become one of the most popular English exams for young learners in China. The *Cambridge English: Key for Schools* (KET) and *Cambridge English: Preliminary for Schools* (PET), designed for EFL learners aged between 11 to 14 years old, have been available in China for over ten years. The *Test of English as a Foreign Language* (TOEFL) has become a major international English proficiency test for Chinese learners who aim to get admission to English-medium universities; according to Qian (2007), Chinese TOEFL test-takers account for about 20% of the total TOEFL population around the globe. The *International English Language Testing System* (IELTS), another famous international English test in China, had over 3 million test-takers in 2010 (G. Yu & Jin, 2016).

On the whole, English language education in China is characterized by an examination-driven culture (Fan, 2018). As a result of such a culture, English teaching and learning has been oriented to the so-called *teaching to tests* or *learning to tests* modes (Chen & May, 2016). Teachers are therefore likely to spend time on test-preparation practices and to place a high value on scores, rather than students' learning progression (Chen, May, Klenowski, & Kettle, 2014; Rao, 2013). Students tend to perceive high scores in tests as the ultimate goal for their learning, which drives them to develop low-level cognitive skills to maximize their scores (Damankesh & Babaii, 2015; Xie, 2015).

Against this background, the Chinese government has made efforts to improve the assessment system by incorporating classroom-based assessment into the original summative assessment system. The primary purpose is to help teachers and students make use of daily assessment information to facilitate teaching and learning, an educational initiative that is consistent with the current worldwide trend towards the use of formative classroom assessment (Black & Wiliam, 1998a; Davison & Leung, 2009). More importantly, this initiative is appropriate for achieving China's educational purpose of "raising the knowledge, skills and well-roundedness ('quality') of a modern citizenry that is globally compatible and future-ready" (Gu, 2014, p. 287).

To sum up, English language education in primary schools in China has expanded rapidly with a growing young EFL learner population in the past two decades. This has contributed to a growing interest in the assessment of young EFL learners in China. The MOE has reinforced the use of classroom-based assessment in the NECS in the hope of improving primary English language learning. Classroom-based assessment principles such as engaging students in the assessment process, identifying clear instructional objectives and success criteria, and using multiple assessment methods are advocated in the NECS. There is little evidence, however, to show how primary EFL teachers conceptualize and implement classroom-based assessment (Yuan & Shu, 2017). Yuan and Shu's review of classroom-based assessment research in the Chinese EFL context identifies that the emphasis has been on tertiary teachers, rather than primary school teachers. Pan and Wu (2019) pointed out that the limited research on CBA in the primary school context has focused on offering advice rather than providing empirical evidence. There is, therefore, an urgent need for research into how primary EFL teachers go about classroom-based assessment in China.

1.3 Theoretical background

For a long time, the assessment field has been dominated by large-scale testing such as school board exams and national standardized tests (Andrade, Bennett, & Cizek, 2019; Lee, Mak, & Yuan, 2019). Since Black and Wiliam's (1998a) landmark research, which demonstrates the crucial role that assessment plays in promoting learning, there has been a paradigm shift to CBA.

A large body of literature provides valuable insights into the principles for using CBA to support learning (e.g., Assessment Reform Group, 2002; Black & Wiliam, 2009; Broadfoot & Black, 2004; McMillan, 2017; Wiliam & Thompson, 2008). To effectively implement classroom-based assessment, it is important to clarify and share clear learning intentions and success criteria with students. Furthermore, the evidence elicited should be instructionally tractable, so that assessments can be used effectively to improve learning. Quality feedback is needed for students to understand the gap between their performance and desired objectives and learn how to close the gap. Students, in particular, need to play an active role in the assessment process.

Some researchers have proposed frameworks for implementing CBA principles in practice (e.g., Buhagiar, 2006; Davison & Leung, 2009; Harlen, 2006b; Rea-Dickins, 2001; Ruiz-Primo, 2011). A core feature of these frameworks is that the implementation of CBA consists of multiple steps. For example, Davison (2008, as cited in Davison & Leung, 2009) proposed a framework of four steps: planning assessment, collecting information about students' learning, making professional judgments, and providing appropriate feedback. Within each of these frameworks, teachers' role is made explicit so that these frameworks can be used as practical guidelines for teachers to implement CBA in their classrooms.

Although there is extensive literature on how CBA can be effective in practice, research into CBA in L2 contexts is relatively limited (e.g., Brumen, Cagran, & Rixon, 2009; Cheng, Rogers, & Hu, 2004; Lee, 2007b; Öz, 2014; Xu, 2016). Nevertheless, the limited research has suggested that L2 teachers do not fully utilize classroom-based assessment to support students' learning. The implementation of CBA is complex as it faces a number of challenges, which relate to teachers, students, and school and educational systems.

Several salient research gaps have been identified in the literature. First, although previous studies have shed some light upon how L2 teachers implement CBA as well as the challenges they face, a comprehensive understanding of L2 teachers' CBA practices is still lacking. A majority of the studies have focused on CBA principles from an individual rather than from a holistic perspective. Given that CBA is a unitary concept (Lee et al., 2019), it is essential to regard CBA principles as interdependent entities. Therefore, a comprehensive investigation of the implementation of CBA is essential.

Second, there is insufficient research on L2 teachers' conceptions of CBA, especially in comparison to the research on teachers' CBA practices. Teachers' conceptions, however, are recognized as a crucial influence on their classroom practices (Borg, 2003). For this reason, it is imperative to explore how teachers conceptualize CBA, and how their conceptions relate to their assessment practices.

Third, the majority of previous studies have been conducted in secondary school and tertiary L2 contexts, with little attention paid to young language learners. As YLL assessment has a unique research agenda, further research is needed to enhance our understanding of the CBA of young learners (Butler, 2016; Hasselgreen, 2012).

This study endeavored to address the above gaps by exploring teachers' conceptions and practices regarding CBA of young language learners in the Chinese context.

1.4 Research questions

The overriding purpose of the current study was to investigate primary school EFL teachers' conceptions and practices regarding classroom-based assessment. Specifically, the study sought to answer the following four research questions:

RQ1: What are EFL teachers' conceptions of classroom-based assessment?

RQ1.1 What are EFL teachers' conceptions of the purposes of classroom-based assessment?

RQ1.2 What are EFL teachers' conceptions of the processes of classroom-based assessment?

RQ2: To what extent do teacher attributes (teaching experience and previous education in assessment) and work environments (school type and grade level) influence their conceptions of classroom-based assessment?

RQ3: How do EFL teachers implement classroom-based assessment practices?

RQ4: What are the relationships between teachers' conceptions and practices regarding classroom-based assessment?

1.5 Significance of the study

It is argued that this study has the potential to contribute theoretically and practically. First, this study adds to the literature on teachers' implementation of classroom-based assessment in L2 contexts. Despite an extensive literature on the principles for

classroom-based assessment, there is limited research on how L2 teachers implement classroom-based assessment. The majority of existing studies have examined limited aspects of CBA. This study, through a comprehensive investigation of the implementation of CBA, provides evidence of how L2 teachers put CBA into practice.

Second, the current study contributes new knowledge to CBA research from the perspective of teacher cognition. Drawing on multiple sources of data, it provides further insights into how primary school EFL teachers conceptualize the purposes and processes of CBA and how their conceptions influence their assessment practices. This study adds to our existing knowledge about how teacher attributes and work environments also influence their conceptions of CBA.

Third, this study adds to the knowledge of the assessment of young language learners. As discussed above, compared with the attention paid to CBA in secondary school and tertiary contexts, little focus has been placed on the CBA of young learners. Assessment of young language learners, as an independent research agenda, has emerged only recently in the field of language assessment (Hasselgreen, 2012). More research has been advocated to investigate every aspect of young language learners, particularly with respect to the benefits of CBA (Butler, 2019). Therefore, this study has the potential to enhance our understanding of CBA for young learners.

Fourth, this study makes a contribution to methodology in providing an example of a mixed-methods approach to examining teachers' conceptions of classroom-based assessment and their assessment practices. Existing studies in this area mainly use questionnaires to elicit L2 teachers' conceptions and self-reported assessment practices, without observing teachers' actual assessment practices. This study demonstrates that a quantitative study to investigate L2 teachers' conceptions and practices, followed by a

qualitative study to look closely at teachers' practices and establish their rationales for such practices, can yield a sophisticated understanding of the relationships between teachers' conceptions and practices.

Finally, this study has practical value in relation to the implementation of classroom-based assessment. It contributes to a comprehensive understanding of the relationships between teachers' conceptions and practices regarding classroom-based assessment. It also identifies factors that potentially hinder teachers from putting their conceptions into actual practices. Thus, this study has implications for policy makers and teacher educators in China, which may also be of relevance to other EFL contexts that share similar conditions.

1.6 Outline of the thesis

This thesis consists of seven chapters. Chapter 1 introduces the thesis and describes the research context and theoretical background, the research questions and the significance of the study.

Chapter 2 presents a review of the relevant literature in three major areas. First, it introduces the origin and definitions of classroom-based assessment, illustrates key principles for CBA as well as a framework of the practice of CBA. The empirical research on CBA in L2 contexts is also reviewed. The second part of this chapter focuses on literature that addresses teacher conceptions, including its origin, definitions and importance, the relationship between teacher conceptions and classroom practice, sources of teacher conceptions, and L2 teachers' conceptions of CBA. The third part reviews the literature on the assessment of young language learners, including the characteristics of young learners and the current status of research on YLL assessment.

Chapter 3 provides an account of the methodology. The research questions are readdressed, followed by a discussion of the philosophical worldview and the overall research design. A detailed description of the procedures of the survey study and the case study is then provided, respectively.

Chapter 4 presents the results of the survey study, which focuses on teachers' conceptions of classroom-based assessment, factors influencing their conceptions, teachers' classroom-based assessment practices, and the relationships between teachers' CBA conceptions and practices. Chapter 5 presents the results of the case study, which illustrates teachers' actual classroom-based assessment practices and rationales for their assessment practices.

Chapter 6 discusses the major findings emerging from this study in relation to the relevant literature. Chapter 7 concludes with a summary of major findings, implications, limitations of the study, as well as recommendations for further research.

CHAPTER TWO

LITERATURE REVIEW

2.1 Chapter overview

This chapter gives a detailed account of the theoretical orientation of this research by reviewing the literature in three areas. The first part of the review focuses on CBA with regard to its origin, definitions, purposes, the principles for using CBA for learning, as well as a framework for the practice of CBA. Empirical studies on CBA in L2 contexts are also reviewed. The second part of the review examines the literature on teachers' conceptions: their origin, definition, and importance. The relationship between teachers' conceptions and classroom practices and the source of teachers' conceptions are discussed, followed by a review of empirical studies on L2 teachers' conceptions of CBA. The third part of the chapter reviews the literature on the assessment of young language learners. The chapter concludes with a summary of the implications drawn from the literature.

2.2 Classroom-based assessment

Teachers' classroom-based assessment offers great promise in supporting students' learning (e.g., Black & Wiliam, 1998a; Brookhart, 2001; Cizek, 2010; Moss & Brookhart, 2009; Panadero, Andrade, & Brookhart, 2018). This part of the review begins by exploring the paradigm shift in the assessment field, a movement from large-scale achievement testing towards teachers' classroom-based assessment.

2.2.1 Paradigm shift in the assessment landscape

Large-scale high-stakes tests, such as school board exams and national standardized tests, are the most "visible and consequential, and (sometimes) controversial tests

encountered in education today” (Cizek, 2010, p. 4). These tests are used primarily for a summative purpose (Abu-Alhija, 2007; Klinger, DeLuca, & Miller, 2008). In other words, these tests are conducted at the end of instruction (e.g., unit, semester, school year), with their primary purpose to yield an accurate and reliable measurement of achievement to be used in decision making such as awarding diplomas or certificates, making promotion, or placing test takers.

Whereas large-scale testing has dominated the field of assessment over the decades, CBA is gaining prominence (Andrade et al., 2019). There are two main reasons for the increasing interest in CBA. For example, although large-scale summative tests have beneficial effects on students’ achievement (e.g., Andrews, Fullilove, & Wong, 2002; Fan, 2018; Phelps, 2005), there is compelling evidence that these tests can exert negative effects on teaching and learning. As some researchers have argued (e.g., Broadfoot & Black, 2004; Cheng, Sun, & Ma, 2015; Pedulla et al., 2003), large-scale tests generally influence teachers to focus on the content areas covered by these tests as well as test-related skills. Students are also likely to focus on test-preparation activities, which may have a detrimental impact on their motivation (e.g., Chik & Besser, 2011; Hamp-Lyons, 2007; Harlen & Crick, 2003; Shepard, Penuel, & Pellegrino, 2018).

There is also a growing body of evidence showing that classroom-based assessment plays a pivotal role in supporting learning (e.g., Andrade & Heritage, 2018; Black & Wiliam, 1998a; Crooks, 1988; Moss & Brookhart, 2009; Natriello, 1987). The following three widely cited literature reviews identify the positive effects of classroom-based assessment on students’ learning. Thirty years ago, Natriello (1987) examined evaluation processes in schools and classrooms. He identified eight elements of the evaluation process that would impact students’ motivation and achievement: the evaluation purposes, the assignment of tasks, clarity of criteria, the setting of standards,

the frequency of sampling learning evidence, the soundness of appraisals, the types of feedback, and the affective value of feedback.

A review by Crooks (1988) focused on the impact of classroom assessment on students' learning strategies, motivation, and achievement. Both formal classroom assessments, such as teacher-made tests and informal curriculum-embedded assessments such as oral questioning and performance activities, were considered. Crooks (1988) concluded that classroom assessment affected students in both direct and indirect ways, and thus deserved careful planning and implementation from educators. Although the reviews by Natriello (1987) and Crooks (1988) have drawn attention to the importance of classroom assessment, they are not comprehensive in scope as only nine studies were cited by both papers.

A decade later, Black and Wiliam (1998a) sought to update the literature on the relationship between classroom assessment and learning by a manual examination of relevant literature on formative classroom assessment in 76 journals published between 1987 and 1997. The literature review included studies from kindergarten to college and covers several disciplines in education. This comprehensive review presents substantial evidence that effective implementation of classroom-based assessment can have profound effects on learning achievement. Moreover, it highlights that formative classroom assessment is particularly helpful to lower-achieving students.

In summary, the assessment field has experienced a paradigm shift in recent decades, with a greater emphasis on classroom-based assessment rather than traditional large-scale testing. Such a shift has led to an increasing focus on the integration between assessment and learning. The next section discusses how the term *classroom-based*

assessment has been defined by researchers and provides a definition of CBA as adopted in this study.

2.2.2 Definitions of classroom-based assessment

The literature shows that the term *classroom-based assessment* has been interpreted by researchers from different perspectives. Broadly speaking, CBA refers to any assessment conducted in classroom settings (e.g., Black & Wiliam, 2018; Davison & Leung, 2009; Leung, 2005; Wang, 2017). This emphasizes that CBA is implemented directly by those who are responsible for teaching. For example, drawing on an analysis of some similar terms, Leung (2005) claimed that CBA is “non-standardized local assessment carried out by teachers in the classroom” (p. 527). Davison and Leung (2009) also pointed out that CBA is “a more teacher-mediated, context-based, classroom-embedded assessment practice, explicitly or implicitly defined in opposition to traditional externally set and assessed large scale formal examinations used primarily for selection and/or accountability purposes” (p. 395). More recently, Black and Wiliam (2018) used the term CBA to describe “those assessments where the main decisions about what gets assessed, how the students will be assessed, and the scoring of the students’ responses, is undertaken by those who are responsible for teaching the same students” (p. 554).

From a technical perspective, CBA is viewed as a set of processes, including collecting evidence of students’ learning, making judgments of the evidence, and using the evidence to make instructional decisions (e.g., Black & Wiliam, 1998b; Davison & Leung, 2009; Harlen, 2006b; Mcmillan, 2013; Turner, 2012). For instance, Zhang and Burry-Stock (2003) defined classroom assessment as “a broad spectrum of activities from constructing paper-pencil tests and performance measures, to grading, interpreting

standardized test scores, communicating test results, and using assessment results in decision-making' (p. 324). This definition clearly suggests a process of evidence elicitation, interpretation, and use. In this respect, CBA is a process-based practice rather than a simple assessment instrument. For example, a test is merely a measurement tool; only when the test is used to elicit evidence of student learning, and relevant instructional decisions are made can the whole process be conceptualized as a CBA activity.

Furthermore, researchers have claimed that classroom-based assessment, as a process-based practice, can be either explicit or implicit (Hill & McNamara, 2011; Mavrommatis, 1997; Rea-Dickins, 2006; Tunstall & Gipps, 1996b). For example, McNamara (2001a) stated that CBA should be “deliberate, sustained and explicit” (p. 343). By comparison, Rea-Dickins (2006) argued that teachers constantly assess their students in ways that are “unplanned and spontaneous (and) inextricably and almost imperceptibly embedded within teacher-learner(s) interactions” (p. 166). Building on Rea-Dickins' (2001) work, Hill (2012) employed the concept of assessment opportunity to account for both the explicit and implicit forms of assessment in the classroom, defining assessment opportunity as:

“...any actions, interactions or artefacts (planned or unplanned, deliberate or unconscious, explicit or embedded) which have the potential to provide information on the qualities of a learner's (or group of learners') performance ”(Hill, 2012, p. 35).

The purpose of CBA, according to some researchers, incorporates both formative and summative purposes (Brookhart, 2004; Hill, 2012; Wang, 2017). For example, Hill (2012) stated that CBA information is used “for teaching, learning, reporting,

management or socialization purposes” (p. 34). Other researchers have used the term to highlight its primary purpose in supporting learning (Davison & Leung, 2009; Leung, 2005; Turner, 2012). According to Turner (2012), the primary purpose of CBA is to “use the information to help make decisions to enhance teaching and learning” (p. 65). Similarly, Rea-Dickins (2007) stated that classroom assessment “has a primarily supportive function in the formative assessment of language learners” (p. 508).

The notion of involving both teachers and learners in the classroom assessment process is well supported by some researchers (Hill, 2012; Stiggins, 2001; Turner, 2012). Hill (2012), for example, characterized CBA as “any reflection by teachers (and/or learners) on the qualities of a learner’s (or group of learners’) work” (p. 34). From a social-constructive perspective, Turner (2012) claimed that teachers and learners are the main participants, with additional participants being parents and school administrators.

Based on a synthesis of the above interpretations, the term CBA is defined, in this study, as:

Classroom-based assessment is an ongoing process (explicit or implicit) that consists of collecting evidence of students’ learning, interpreting and using such evidence; the process involves both teachers and learners and serves a primary purpose of supporting learning.

The next section provides a description of how researchers in the assessment field have conceptualized the purposes of classroom-based assessment.

2.2.3 Purposes of classroom-based assessment

Much of the discussion on the purposes of classroom-based assessment has focused on the distinction between summative and formative. The distinction was originally related to the ‘timing’ aspect of assessment. The concepts of formative and summative

evaluation were first coined by Scriven (1967) in a monograph of the American Educational Research Association (AERA). According to Scriven (1967), timing was the primary distinction between the two types of evaluation. Specifically, formative evaluation was conducted *during* a school program, whereas summative evaluation was conducted *after* completion of the program. Note that the concepts put forward by Scriven (1967) are focused upon the evaluation of school program rather than student learning.

Scriven's (1967) concepts of formative and summative evaluation were first applied to the process of student learning in the early 1970s when an influential volume was published by Bloom and his colleagues, entitled the *Handbook of Formative and Summative Evaluation of Student Learning* (Bloom, Hastings, & Madaus, 1971). In this handbook, timing was still characterized as an essential distinction between formative and summative assessment. Bloom et al. (1971) stated that summative assessment occurred after the instruction had taken place, whereas formative assessment took place during learning and focused on smaller units of instruction. The most significant contribution, however, made by Bloom and his colleagues is that they suggested a link between the formative approach to assessment and the improvement of learning. According to Bloom et al. (1971):

“Formative evaluation is for the use of systematic evaluation in the process of curriculum, construction, teaching, and learning for the purpose of improving any of these three processes” (p. 117).

Towards the end of the 1980s, the distinction between summative and formative assessment turned to the *purpose* of assessment. According to Sadler (1989), the essence of formative assessment was the use of results to shape and improve students'

competence. In contrast, summative assessment was primarily concerned with summarizing and reporting students' achievement; it did not have an immediate impact on learning. He emphasized that the primary distinction was related to purpose and effect rather than timing.

Currently, many researchers have distinguished formative and summative assessment in terms of the purpose of assessment. Black (1998), for instance, has defined formative as “to aid learning” and summative as “for review, transfer and certification” (p. 34). Harlen (2006b) has also made a point of emphasizing that “what lies at the heart of the distinction between formative and summative is not the timing and frequency, but the use of evidence, who uses it, how and for what purposes” (p. 29). In the same vein, Black and Wiliam (2018) have stated that the distinction is related to “the kinds of references being drawn from assessment outcomes” (p. 553). They have argued that an assessment serves a formative function if the inferences concern what kinds of actions will best support students to learn, and a summative function if the references are about the status of students' learning.

Some researchers have conceptualized formative assessment and summative assessment as opposite ends of a continuum (e.g., Dixson & Worrell, 2016; Harlen, 2006b; Harlen & James, 1997; Rea-Dickins, 2007). From their perspective, the same assessment instruments, even the same assessment results, can be used to serve both formative and summative purposes. For example, Black and Wiliam (1998b) have stated that summative tests can be used to serve the formative function of supporting learning. Rea-Dickins (2007) has stressed that the boundary between formative and summative assessment is not as clear cut as usually represented because teachers may use the same assessment information for different purposes at different times. For instance, a

language sample of a child can be used formatively to inform teachers' planning, while the same sample can be used to record the child's language achievement.

While the distinction between formative and summative assessment has attracted much attention, a number of researchers have tried to understand also the nature of formative assessment. Earlier interpretations of formative assessment highlighted the use of assessment to adjust instructional activities (including both teaching and learning) to better meet students' needs. A representative definition is one proposed by Black and Wiliam (1998a):

“Encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (p. 7).

Although this definition does not describe what kind of assessment is formative, it provides a basis for many interpretations. Some researchers have proposed more narrowed definitions of formative assessment by requiring that the instructional adjustments should take place during the instruction. Cowie and Bell (1999), for instance, defined formative assessment as “the process used by teachers and students to recognize and respond to student learning to enhance that learning, during the learning” (p. 32). Similarly, Looney (2005) emphasized that formative assessment should be conducted while learning was taking place. He defined formative assessment as a “midstream” tool, which was used to “identify specific student misconceptions and mistakes while the material is being taught” (Looney, 2005, p. 11).

Some researchers have argued that these earlier interpretations are less likely to highlight the power of formative assessment in supporting learning (Assessment Reform Group, 1999; Gardner, 2012; Stiggins, 2005). The Assessment Reform Group (1999)

argued that the term *formative* is open to a number of interpretations, and has often been misunderstood as a process in which frequent assessments are carried out during the instruction over time and are collated for a summing-up purpose. Gardner (2012) noted that this interpretation equates formative assessment to “the summative use of multiple assessments of learning” (p. 2). Whilst such assessments may be useful in helping teachers identify areas for improvement and telling students about their success or failure, they do not provide information about the next steps towards further learning.

To highlight the power of assessment in facilitating learning, the Assessment Reform Group recommended the term *assessment for learning* (AfL). They defined AfL as the “the process of seeking and interpreting evidence for use by learners and their teachers, to identify where the learner are in their learning, where they need to go and how best to get there” (Assessment Reform Group, 2002, pp. 2–3). As this definition implies, the core of effective assessment is the promotion of assessment to support learning. The most important feature of this definition may be that it describes the key characteristics of assessment that are formative. It draws attention, specifically, to three instructional processes: identifying where the learners are in their learning, identifying where learners need to go, and identifying how learners can best get there. They also argued that, to maximize the impact of assessment on learning, students should be assigned a role in the assessment process.

A more comprehensive recent definition of AfL was put forward during the Third Conference on Assessment for Learning in 2009. AfL has been defined as “part of everyday practices by students, teachers, and peers that seek, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning” (Klenowski, 2009, p. 264). The phrase “everyday practice” emphasizes that assessment is an integral part of teaching and learning. The phrase “by

students, teachers and peers” asserts that AfL should be student-centered because all assessment practices by teachers can eventually be effective only when students can take necessary actions to improve. The phrase “seeks, reflects and responds to” indicates that AfL works as a set of processes of collecting evidence of student learning, interpreting evidence, and using such evidence for wise decisions for the next steps. The phrase “information from dialogue, demonstration and observation” emphasizes that both planned and unplanned events can be used as sources of evidence. The phrase “in ways that enhance ongoing learning” confirms that assessment can be formative only when students and teachers use the information to enhance learning. This definition makes it clear that the primary purpose of classroom-based assessment is the learning improvement on the part of students.

Black and Wiliam (2009) also proposed the following comprehensive definition of formative assessment:

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

As this definition implies, the essence of formative assessment lies in the extent to which assessment enhances teaching and learning. Another important feature is that it addresses the fundamental characteristics of formative assessment, including making decisions about the next steps and encouraging learners to be engaged in their learning.

To sum up, the distinction between summative and formative has changed from a focus on the *timing* to a focus on the *purpose* of assessment. The meaning of formative

assessment has also evolved. The earlier interpretations, focused on making instructional adjustments, did not imply that assessment had a role in supporting learning. Current interpretations are concerned with the extent to which evidence is used to regulate the learning process.

2.2.4 Principles of classroom-based assessment for learning

Given that the primary purpose of CBA is to support students' learning, many researchers have attempted to identify the principles that underpin effective CBA practices (e.g., Assessment Reform Group, 2002; Black & Wiliam, 2009; Lee, 2007a; Sadler, 1989; Shepard, 2006). Based on the synthesis of the literature, four key principles are discussed in this review: articulation of learning intentions and success criteria, evidence elicitation, feedback provision and student involvement.

2.2.4.1 Articulation of learning intentions and success criteria

Clear learning intentions are critical to the integration of teaching, learning and assessment (Chappuis, Stiggins, Chappuis, & Arter, 2012; Panadero et al., 2018; Ruiz-Primo, 2011; Russell & Airasian, 2012; Wiliam, 2007) As argued by Ruiz-Primo (2011), clear learning intentions help teachers plan appropriate instructional activities, and select assessments to determine whether students have achieved the intended knowledge and skills. Sharing learning intentions with students enables them to clarify their learning goals, thus enabling them to be truly engaged in the learning process (Brookhart, Moss, & Long, 2008; Sadler, 1989). As Timperley and Parr (2009) claimed, without a clear understanding of what they are expected to achieve, students are likely to focus on the surface features of a task, believing that their ultimate goal is to complete the task. Moreover, clear learning intentions give students opportunities to assess and self-monitor their own learning (Glasson, 2009).

The literature on CBA draws attention to the distinction between mastery goals and performance goals (Ames, 1992; Dweck, 1996, 2000; Hawe & Parr, 2014). Mastery goals focus on “the knowledge, behavior, skill or strategy students are to acquire” (Schunk, 2000, p. 36). These goals encourage students to understand the nature and purpose of an activity, to have a willingness to use feedback to improve learning and to persist when they face difficulties. In contrast, if the goal is performance in orientation, it emphasizes “demonstrating competence, being superior to others, and the use of social comparative or normative standards” (Pintrich, Conley, & Kempler, 2003, p. 321). As a result, students are motivated to seek the easiest way to gain higher achievement rather than to develop general life goals (Harlen, 2006c), which can result in low self-efficacy among low-achieving students (Schunk, 2000). For this reason, many researchers have emphasized that students should be oriented towards learning rather than performance (e.g., López-Pastor & Sicilia-Camacho, 2017; McMillan, 2010; Torrance & Pryor, 2001). For instance, Torrance and Pryor (2001) stated that teachers need to focus students’ awareness of “the intended learning goals of the activity, rather than on performance goals” (p. 625).

Teachers also need to take into account students’ learning needs (Airasian, 2012; Hattie & Jaeger, 1998; McMillan, 2011; Sadler, 1989) to set goals that are achievable for students. From a theoretical perspective, this reflects Vygotsky’s (1978) zone of proximal development (ZPD) theory, in which the ZPD is the difference between “the actual developmental level as determined by independent problem solving and the level of potential development as determined” (p. 86). When applying the theory to the process of student learning, teachers should build on students’ current understandings and facilitate learning within their ZPD. In this sense, students’ current needs must be considered in order to set suitable learning goals.

To promote a classroom environment that is conducive to learning, it is also crucial to make success criteria explicit to students (e.g., Andrade & Brookhart, 2016; Clarke, 2005; Glasson, 2009; McMillan, 2011a). Success criteria refers to the “facets or dimensions of performance that are used for judging the level of achievement” (McMillan, 2011, p. 34). According to Clarke (2005), sharing success criteria enables students to “apply appropriate focus, clarify understanding, identify success, determine difficulties, discuss strategies for improvement (and) reflect on overall progress” (p. 37). In other words, when students have a clear idea of success criteria, they are more likely to take on responsibility for their own learning process. White and Frederiksen (1998) reported that students could improve their scores when there were recognizable criteria for good work. They found that communicating success criteria to students with weak basic skills was particularly important because, if they did not understand the criteria, they were unlikely to achieve their learning goals.

Giving success criteria to students makes little sense unless they understand what the criteria mean (Elwood & Klenowski, 2010; López-Pastor & Sicilia-Camacho, 2017; Sadler, 1989). Teachers can help students understand the success criteria by giving them exemplars of strong or weak work and by involving students in using the criteria to evaluate the exemplars (Chappuis, 2005). In doing so, students can familiarize themselves with the success criteria as well as develop necessary self-assessment skills (William, 2007). Another way of helping students understand the criteria is to invite students to discuss what might be the criteria for the quality of a piece of work (Clarke, Timperley, & Hattie, 2001).

2.2.4.2 Evidence elicitation

A variety of methods or procedures can be used to elicit information about students' understanding (Chappuis et al., 2012; Cizek, 2010; Wiliam, 2010a). As noted by Wiliam (2010a), while evidence elicitation often takes the form of teacher questioning, any actions that provide information on students' learning can also be included. McMillan (2008) has categorized these methods into two major types, namely traditional assessments, and alternative assessments. Such a categorization is based on the link between assessment and instruction. Traditional assessments refer to those that are constructed to measure students' performance with paper-and-pencil tests, including multiple-choice tests, sentence completion, and essays. In contrast, alternative assessments are used to make students' learning more explicit and to create a closer link between assessment and instruction. Types of alternative assessments include authentic assessment, performance assessment, portfolio assessment, student self-assessment, and peer assessment.

Diverse assessment formats, however, do not elicit evidence that is equally useful. As noted by Wiliam (2010a), certain kinds of evidence serve only monitoring (signal which students have difficulties) or diagnostic (locate the specific difficulties) function. The most useful assessments are those that yield information that is instructionally tractable. In other words, such assessments not only yield information that indicates if there is a problem, and what precisely the problem is, but also provide information on which to base the next steps for moving students forward. A challenge, therefore, is to ensure that the evidence elicited is instructionally tractable so that assessments can be used to improve learning.

For evidence to be instructionally tractable, assessments, first, must provide a sufficient level of detailed information about students' learning (Wiliam, 2010a). More importantly, such detailed information needs to be actionable (Heritage, 2013). That is, assessments need to reveal students' thinking to provide insights into what students need to do to progress.

2.2.4.3 Feedback provision

In general, feedback refers to “information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding” (Hattie & Timperley, 2007, p. 81). It can be provided in different forms, such as teachers' written comments or oral responses. Regardless of its form, feedback plays a pivotal role in classroom-based assessment (Black, 2015; Black & Wiliam, 2009; Brown, Peterson, & Yao, 2016; Cowie & Bell, 1999; Wiliam & Leahy, 2007; Yoon & Lee, 2013). As suggested by Black and Wiliam (2009), any assessment can be formative if the evidence generated is used as feedback to make better instructional decisions than those made without such kind of evidence. Feedback, thus, is crucial to the improvement of the instructional process.

A defining feature of feedback, therefore, is that it must exert some consequences on the instructional system. This is emphasized by Ramaprasad (1983), who defined feedback as “information about the gap between the actual level and the referenced level of a system parameter which is used to alter the gap in some way” (p. 4). Ramaprasad (1983) claimed that information that fails to fulfill its formative function of changing the instructional system is not feedback. Sadler (1989) also noted that the information provided could be regarded as feedback only when it alters the gap between students' current and desired performance. Sadler (1989) argued:

“If the information is simply recorded, possessed to a third party who lacks either the knowledge or the power to change the outcome, or is too deeply coded (for example, as a summary grade given by the teacher) to lead to appropriate action, the control loop cannot be closed, and “dangling data” substituted for effective feedback” (p. 121).

For feedback to be effective in advancing learning, attention should be paid to the quality of feedback (Black & Wiliam, 1998a; Brookhart, 2008; Hattie & Timperley, 2007; Sadler, 1989). According to Black and Wiliam (1998b), quality feedback should be concerned with the quality of students’ work and provide specific advice on what students can do to improve their learning. Wiliam (2011) reiterated this, stating that good feedback practice is more than information about the disparity between current and desired performance, it also provides insights into the kinds of instructional activities that can be undertaken to improve performance. Likewise, Sadler (1989) argued that feedback must satisfy three conditions if it is intended to be effective: communicating the standards, comparing the current level of performance with the standards, and engaging students in actions that are likely to result in closure reduction of the learning gap. Hattie and Timperley (2007) summarized the three conditions as three questions: ‘Where am I going?’, ‘How am I going?’, and ‘Where to go next?’. They proposed that the focus of feedback can be categorized into four levels: (1) feedback about the task (focusing on how well a task is being accomplished); (2) feedback about the processing of the task (focusing on information about how students complete a task); (3) feedback about self-regulation (focusing on the strategies that students use to monitor, control, direct or regulate their actions towards the learning goal); and (4) feedback about the self as a person (focusing on the personalities of students, which is unrelated to the task performance). They argued that feedback is most

effective if it is directed at the first three levels. In contrast, feedback on the self-level is less effective as it contains little task-related information that provides answers to the three feedback questions, and thus contributes little to learning.

2.2.4.4 Student involvement

The early conceptualizations of classroom-based assessment put great emphasis on the teachers' role (Harlen & James, 1997; Stiggins & Conklin, 1992; Tunstall & Gipps, 1996b). Stiggins and Conklin (1992), for instance, defined classroom assessment as "those assessments developed and used by teachers in their classroom on a day-to-day basis" (p. 1). They profiled eight dimensions of teachers' classroom assessment environment, including teachers' assessment purposes, the assessment methods used, the criteria for selecting the methods, the quality of the methods, teachers' feedback, teachers' preparation for being an assessor, teacher' perceptions of students, and assessment policy environment. This conceptualization highlights that classroom assessment environment was shaped by teachers' approach to assessment because seven of the dimensions were directly related to teachers' role, with an assessment-policy environment being the exception.

Early conceptualizations have been criticized for overemphasizing teachers' role and underplaying students' role in the assessment process. For example, Perrenoud (1998) has pointed out that the discourse of classroom assessment is "solely concerned with formative evaluation practices of teachers" (p. 99). Brookhart (1997) has argued that the framework of classroom assessment environment proposed by Stiggins and Conklin (1992) reflects the instructional processes but fails to respect students' place in the environment.

Researchers have noted that the overemphasis of teachers' role reflects the behaviorism learning theory (Pryor & Crossouard, 2008; Torrance & Pryor, 1998). Behaviorism learning theory emphasizes that learning is a result of stimulus and response (Budiman, 2017; Shepard, 2000). Behaviorists posit that learners are passive receivers whose primary task is to respond to environmental stimuli (Gagné, 1965; Kay & Kibble, 2016). In contrast, the teachers' role is to design and ensure that students respond to the environment stimuli correctly and appropriately (Fauziati, 2016; James, 2006). Shaped by behaviorism learning theory, it is taken for granted that teachers are the authoritative figure and should play a dominant role in the process of assessment (Torrance & Pryor, 1998).

Recently, socio-cultural learning theory has received a great deal of attention. Socio-cultural theories of learning posit that learning occurs in a social and collaborative environment (Brookhart, 2018; James, 2006; Ormrod, 2014). Through a sociocultural lens, students are perceived as active agents who construct knowledge through social interactions rather than passive individuals who receive and memorize information (John-Steiner & Mahn, 1996; Mitchell, Myles, & Marsden, 2013; Packer & Goicoechea, 2000). Teachers are not perceived as the sole authoritative source of knowledge; instead, their role is to assist novice learners in constructing knowledge and in moving towards self-regulation of learning. Influenced by the sociocultural learning theory, classroom assessment is now conceptualized as an interaction between teachers and students, with students playing a pivotal role in the process of assessment (Allal, 2016; Elwood & Klenowski, 2010; Panadero, Broadbent, Boud, & Lodge, 2019; Pryor & Crossouard, 2008).

The involvement of students has been explicitly incorporated in the current notions of classroom-based assessment. For example, the Assessment Reform Group

(2002) proposed the concept of AfL, in which assessment was described as “a process of seeking and interpreting evidence for use by learners and their teachers to decide where learners are in their learning, where they need to go and how best to get there” (p. 2). Wiliam and Thompson (2008) similarly emphasized that formative classroom assessment could be viewed as encompassing three key processes (identifying where learners are going, where they are in their learning, how they can close the learning gap), and the processes are undertaken by three groups of actors (teachers, learners, peers).

Self-and peer assessment has been seen as a core aspect of classroom-based assessment to promote self-regulated learning (Andrade, 2018; Boud & Falchikov, 1989; Glasson, 2009; Panadero, Jonsson, & Botella, 2017). Self-assessment refers to a “descriptive and evaluative act carried out by the student concerning his or her own work and academic abilities” (Brown, 2013, p. 368). According to this definition, self-assessment involves students in using success criteria to evaluate and monitor their own learning. If self-assessment is conducted effectively, it can develop students’ metacognition skills, improve their self-regulation skills and help them become lifelong learners (Elwood & Klenowski, 2010; Panadero et al., 2018; Ramdass & Zimmerman, 2008).

Peer assessment requires students to make judgments of the work of their peers based on pre-determined criteria and to provide feedback to their peers (Falchikov, 1995; Reinholz, 2016; Wanner & Palmer, 2018). It provides students with opportunities to reflect on their learning, generate inferences, and explain their understandings (Andrade & Brookhart, 2016; Black, 2015; Falchikov, 1995). Thus, peer assessment has a variety of benefits for students, such as improving students’ responsibility for their learning and creating a collaborative learning environment (Falchikov, 2005; Sadler, 1989).

Researchers have identified several conditions for effective self-and peer assessment. First, students must have a clear understanding of the success criteria because it should be used to evaluate their own work (Elwood & Klenowski, 2010; Panadero, Jasson, & Strijbos, 2016). It is recommended that students should be involved in reviewing models of self-and peer assessment or co-establishing the criteria (Andrade, 2010; Topping, 2003). Second, students must be aware of the value of self-and peer assessment (Goodrich, 1996). If students do not have an awareness of the benefits of self-evaluation, they cannot be fully engaged in self-assessing their own or peers' work as it requires their efforts (Panadero & Alonso-Tapia, 2013). Further, teachers should teach students how to use specific criteria to assess their own work and also to provide feedback to students about their assessment process so that they can understand whether it has been implemented appropriately (Panadero et al., 2017; Ross, 2006; Topping, 2013). In addition, it is important to create a supportive classroom culture in which students feel comfortable expressing their opinions (Glasson, 2009).

In summary, this section has provided a description of four fundamental principles that are critical for formative CBA. The next section presents a straightforward framework for the practice of CBA, illustrating how teachers can be instructed to implement CBA principles effectively in their classrooms.

2.2.5 A framework for the practice of CBA

The implementation of CBA is often described as a process with multiple steps (e.g., Buhagiar, 2006; Davison & Leung, 2009; Harlen, 2006b; Rea-Dickins, 2001; Ruiz-Primo, 2011). This study is based on Davison's framework (2008, as cited in Davison & Leung, 2009), in which teachers' assessment is viewed as a cyclical process that consists of four steps: planning assessment, collecting information about students'

learning, making professional judgments, and providing appropriate feedback or advice. In the following subsections, these four steps are described from the perspective of teachers' effective implementation of CBA principles in practice.

2.2.5.1 Planning assessment

Planning assessment is the fundamental step of classroom-based assessment (Lee, 2007b; Xu, 2016). It starts with clarifying the learning goals or objectives, as well as the success criteria that determine what will be recognized as evidence of achieving these goals. As mentioned earlier, teachers should promote mastery goals that emphasize the understanding and mastery of knowledge and skills rather than performance goals that emphasize the comparison of students' abilities (Cauley & McMillan, 2010). It is necessary for teachers to establish clear learning goals that represent national or state standards (Abedi, 2010). Since state standards often define the knowledge and skills that students are expected to learn at each grade level, assessment that is aligned with the state standards will provide meaningful information about "whether students are making adequate progress toward achieving the standard" (Herman & Baker, 2005, p. 2). This requires teachers to deconstruct state standards into achievable learning objectives that students must master on their journey to meet the standards. Furthermore, teachers should take students' learning needs and backgrounds into account to set goals that are achievable for students (Sadler, 1989). Teachers can explain the learning goals and the success criteria to students using student-friendly language or visual images (Clarke et al., 2001), or engage students in discussing the learning goals and developing the success criteria (Ruiz-Primo, 2011). Teachers can also help students understand the success criteria by giving them exemplars of strong or weak work and by supporting students to apply the criteria to evaluate the exemplars (Chappuis, 2005).

Selecting appropriate assessment methods is also an important component of planning assessment. Although a number of different types of assessment methods are available, the driving force behind the selection of an appropriate method is to match an assessment method to an intended objective (Rea-Dickins, 2001). It is also important that teachers know the strengths of various assessment methods so that they can select an appropriate method that suits the instructional objective (McMillan, 2011).

2.2.5.2 Collecting evidence about students' learning

The next step is collecting revealing and pertinent evidence about students' learning achievement, evidence that is instructionally tractable (Heritage, 2013; Wiliam, 2010a). As discussed earlier, the evidence collected should indicate where students are in relation to the learning intentions as well as provide insights into students' thinking (Heritage, 2013). The wide range of assessment methods is often understood as a continuum determined by the spontaneity of the assessment moment, with informal assessment at one end and formal assessment at the other (Araceli Ruiz-Primo & Furtak, 2006; Hill & McNamara, 2011; McMillan, 2011; Turner & Purpura, 2016). According to Hill and McNamara's (2011) taxonomy, there are three types of assessment methods: (1) spontaneous assessment opportunity; (2) planned assessment opportunity; and (3) formal assessment. This section discusses these methods from the perspective of assessment for learning.

In classrooms, teachers frequently use spontaneous assessment opportunities during any teacher-student interaction or instructional activity. As suggested by Bell and Cowie (2001), these assessment opportunities are "embedded and strongly linked to learning and teaching activities" (p. 86). The timeframe for evidence interpretation is immediate, and teachers' response to the evidence is usually spontaneous. According to

Turner and Purpura (2016), most incidental assessment opportunities are generated by teachers to help them notice, understand, and use learning evidence.

Two major types of spontaneous assessment opportunities are questioning and observation. Questioning is a powerful strategy for formative classroom-based assessment (Assessment Reform Group, 1999; Black & Wiliam, 1998a). To be effective in supporting learning, open-ended questions that probe insights into students' thoughts should be used (Ruiz-Primo, 2011). In doing so, deep learning can be advanced. Closed questions that focus on factual recall or a particular response may result in superficial learning (Harlen, 2007). Observation engages teachers in observing students' learning performance through incidental happenings, such as what students are saying or doing (Airasian & Abrams, 2003). Through observations, teachers can obtain contextualized and authentic information about students' learning (Baker, 2006).

Planned assessment opportunities involve the use of regular instructional activities. Such assessment opportunities are fully embedded in instruction and typically start with students doing an activity, with teachers then immediately reflecting on and acting on the evidence collected (Ruiz-Primo & Furtak, 2006). The instructional activities are often in the form of projects, class discussions, and group work, and used mainly to generate information about students' learning at checkpoints to inform ongoing instruction. In some cases, students are actively involved in the assessment process (Turner & Purpura, 2016). In L2 contexts, for example, they will be asked to use the words they have learned to produce a piece of work. They are given the success criteria and asked to evaluate their own work, or provide feedback to their peers, which is expected to promote self-regulated learning.

Formal assessment takes place when more traditional measures such as tests and homework are used to provide evidence of learning achievement, although the evidence is also used to provide feedback (McMillan, 2011; Turner & Purpura, 2016). In other words, this type of assessment is essentially summative, primarily employed to evaluate students' performance and assign grades, but with the added purpose of improving learning. As claimed by Turner and Purpura (2016), the information gathered through formal assessments can be used to “infer both where students are in their development and the kind of scaffolding assistance they need to further processing and to close the learning gaps” (p. 263). Formal assessments often take the form of teacher-designed tests, homework, quizzes, as well as unit or midterm tests.

2.2.5.3 Make professional judgments

Making professional judgments involves making sense of the evidence collected. Researchers have argued that it is imperative for teachers to interpret students' responses in relation to established standards, which is criterion-referenced assessment (Cumming, 2009; Harlen & James, 1997; Lok, McNaught, & Young, 2016). When criterion-referenced assessment is used, the judgment of students' work is made in terms of how students' performance matches the criteria describing specific skills, ideas or knowledge; the judgment does not depend on how other students perform (Harlen, 2006b). Criterion-referenced assessment helps diagnose the specific nature of any problems in learning and identify the action that needs to be taken (Harlen & James, 1997; McMillan, 2008). It is claimed to be an effective way to strengthen students' confidence and motivation and to overcome the undesirable negative effects of comparison (Airasian & Abrams, 2003; Black, 2001).

Harlen (2006a) argued that ipsative assessment, or pupil-referenced assessment, that is when the judgment takes into account the student's efforts and the progress that the student has made over time, must be considered if assessment is to improve learning. Harlen and James (1997) noted that if evidence interpretation is purely criterion-referenced, it can discourage students who often have difficulty in achieving the desired standards. According to Jacobs and Renandya (2019), pupil-referenced assessment can promote self-directed learning as students can diagnose their own learning needs, as well as identify the resources and strategies to help achieve their learning objectives.

Norm-referenced assessment, in contrast, is seen as less helpful than criterion-referenced and pupil-referenced assessment. When norm-referenced assessment is used, the judgment is made by comparing each student's performance with that of their peers (Harlen, 2006a). Although norm-referenced assessment may help determine whether students have demonstrated average knowledge, it fails to identify the specific nature of the knowledge and thus offers no help in terms of what to do to close a learning gap (Harlen & James, 1997; McMillan, 2008). In addition, norm-referenced assessment emphasizes scores rather than students' efforts, which can have a detrimental impact on students' attitudes to learning and motivation (Berry, 2011; Darandari & Murphy, 2013; Lok et al., 2016). As assessment of students' success is dependent largely on others' performance, it may also exert a negative impact on cooperation among students (Boud & Falchikov, 2007; Jacobs & Renandya, 2019).

Making professional judgments requires the interpretation of evidence to be valid and reliable. Validity refers to the soundness or trustworthiness of interpretation made on the basis of obtained information (McMillan, 2011); it can be determined using: (1) content-related evidence (i.e., adequate sampling of content or knowledge to be assessed); (2) criterion-related evidence (i.e., an assessment provides the same results

as another assessment of the same thing); and (3) consequential evidence (i.e., intended and unintended effects on teaching and learning). Reliability is concerned with the dependability of the information observed (Russell & Airasian, 2012).

2.2.5.4 Providing appropriate feedback

Feedback should empower students to understand the disparity between current and desired performance and the kinds of actions that can be undertaken to improve performance (Brookhart, 2008; Wiliam, 2011). To effectively promote learning, teachers should provide descriptive, rather than evaluative feedback, to students (Nicol & MacFarlane-Dick, 2006; Tunstall & Gipps, 1996a; Wiliam, 2010a). According to Tunstall and Gipps' (1996) feedback typology, descriptive feedback includes strategies of *specifying attainment* (e.g., teachers acknowledge specific components of attainments), *specifying improvement* (e.g., teachers specify how something can be corrected), *constructing achievement* (e.g., teachers draw students into articulating or demonstrating achievement) and *constructing progress* (e.g., teachers give students the responsibility to make choices for their own learning). When descriptive feedback is used, it can lead students to recognize the gap between actual and desired learning goals, and to be able to monitor their own learning to close the gap (Gamlem & Smith, 2013; Glasson, 2009; Wiggins, 2012).

Evaluative feedback, in contrast, often comes in the form of *rewards* (e.g., stickers, smiley faces, stars, symbols, marks, letter grades), *punishments* (e.g., removing the rewards or assigning extra homework), expressions of *approvals* (e.g., smiling, making eye contact, making general comments like 'Well done') and *disapprovals* (e.g., criticizing students) (Tunstall & Gipps, 1996b). While this kind of feedback tells students whether they have performed a particular task correctly or not, it provides them

with little information about how to improve their learning (Murtagh, 2014; Nicol & MacFarlane-Dick, 2006). As evaluative feedback focuses on comparisons, it can impact negatively on students' self-efficacy (Brookhart, 2001).

In summary, the discussions above illustrate the four steps of CBA from the perspective of what teachers can do to effectively promote learning. This framework was regarded as important to the current study for two reasons. First, as stated by Davison and Leung (2009), this framework is informed by the assessment for learning philosophy. The multiple steps correspond to three key formative assessment questions, which are summarized by Ramaprasad (1983) as: (1) Where am I going? (2) Where am I now? (3) How can I close the gap? The *Where am I going?* question addresses the issue of clarifying and sharing clear learning intentions and success criteria with students. The *Where am I now?* question refers to the specific activities in which evidence of learning is elicited and interpreted. The *How can I close the gap?* question suggests the practice of providing feedback that moves students forward. The framework also draws attention to the role that students play in the assessment process. That is, the big idea that integrates the multiple steps is that learning evidence is used with an overriding aim of helping students become self-regulated learners.

Second, teachers' role, central to these four steps, is made explicit in this framework. The framework, therefore, provides practical guidance for teachers implementing CBA effectively during their daily instruction. For this reason, this framework was used in this study to understand teachers' understandings and practices in relation to CBA.

While the implementation of CBA is described as a process with multiple steps, it does not mean that assessment necessarily proceeds through these steps as in a 'linear'

way. As Buhagiar (2006) reported, the multiple steps are interactive with blurred distinctions between them. Likewise, Harlen (2006b) emphasized that these steps are not ‘fixed’ but represent a framework for understanding what teachers can do to promote the ‘assessment for learning’ philosophy.

Having discussed the principles that inform effective CBA practices and presented a framework for the implementation of CBA at the classroom level, the next section reviews empirical research on CBA in L2 contexts.

2.2.6 Research on classroom-based assessment in L2 contexts

Although there is extensive literature on how classroom-based assessment can be used to promote learning in general education, empirical research on CBA in L2 contexts is relatively rare. The limited research, nonetheless, provides insights into the current status and challenges for L2 teachers’ CBA implementation.

2.2.6.1 Current status of L2 teachers’ CBA implementation

Current research has shown that CBA has not been fully implemented to support students’ learning in L2 contexts (Butler, 2009; Büyükkarcı, 2014; Chen et al., 2014; Xu, 2016). Specifically, teachers seem to place little emphasis on the practices of sharing learning intentions and success criteria with students. When collecting evidence, it seems that multiple assessment methods are used, but with a reliance on more traditional assessment methods rather than alternative assessment methods that link assessment with instruction. There is also evidence that teachers often provide evaluative or judgmental feedback, with little focus on descriptive feedback that moves students forward. Furthermore, teachers seem to play a dominant role in the assessment process, regardless of students’ agency in learning. A review of research into L2 teachers’ CBA practices is presented in the following subsections.

Teachers' practices of sharing learning intentions and success criteria

There is evidence that L2 teachers place little emphasis on clarifying learning intentions (Cheng & Wang, 2007; Gu, 2014; Xu, 2016; Zhou & Deneen, 2016). Gu (2014), for instance, investigating the CBA practices of an EFL teacher in a Chinese secondary school, reported that the teacher seldom set specific learning expectations for students and that the teaching practice was entirely textbook-bound with a focus on students' performance goals. In another study, Xu (2016) investigating the assessment planning practices of 20 university EFL teachers in China, reported that while the teachers were conscious of developing the teaching objectives at the beginning of the lesson plans, they did not align the objectives with their instructional activities.

Other studies have demonstrated that L2 teachers tend not to make the criteria for success explicit to students. For example, Zhou and Deneen (2016), in a study of Chinese tertiary teachers' self-reported classroom-based assessment practices, found that the teachers did not make achievement-related criteria transparent to students prior to task completion. Cheng and Wang's (2007) interview study explored teachers' classroom assessment practices at the tertiary level in three ESL/EFL contexts, Canada, Hong Kong, and China. Although the majority of the teachers in the three contexts prepared scoring criteria prior to assessment tasks, there was variation in their practices of sharing the criteria with students. Although most teachers in Canada and Hong Kong informed their students of the criteria by providing explanations or demonstrations, few Chinese teachers attempted to clarify the criteria.

Teachers' practices of collecting learning evidence

Studies have suggested that L2 teachers use multiple assessment methods to collect evidence of student learning (Brumen et al., 2009; Cheng et al., 2004; Cheng & Sun,

2015; Gan, Leung, He, & Nang, 2018; Yang, 2012). For instance, Yang (2012), in a survey of 27 EFL teachers at tertiary level in Taiwan China, found that teachers employed both traditional assessments such as multiple-choice questions and true/false questions, and alternative assessment such as informal questioning, observations, teacher-student conferences, self-assessment, and peer assessment. Gan et al. (2018), in a survey of 204 EFL teachers at tertiary level in China, found that the teachers used a variety of assessments, which included in-class formal assessment (e.g., reading aloud, dictation, quizzes and textbook exercises), interactive informal assessment (e.g., oral questioning, observation), performance assessment (e.g., essay writing, translation) and self-assessment. Likewise, Cheng et al. (2004) provided evidence that tertiary ESL/EFL teachers from Canada, Hong Kong, and Beijing employed three categories of assessment methods: instructor-made assessment methods (e.g., short answer items, true/false items, multiple-choice items), student-conducted assessment methods (e.g., student journal, peer assessment, oral questioning, self-assessment, portfolio), and standardized testing. Thus, there is considerable evidence that teachers use a variety of assessment methods to collect learning evidence in L2 classrooms, although how they are categorized differs among the researchers.

As well as indicating the use of multiple assessment methods, the existing literature reports variation in the frequency different methods are used (Brumen et al., 2009; Cheng, 2015; Öz, 2014; Saito & Inoi, 2017). In general, L2 teachers rely heavily on traditional and formal assessments such as multiple-choice items, essays, and quizzes; and alternative assessments, such as oral questions, students' self-and peer assessment, portfolios, are less frequently used. For example, Öz's (2014) study of Turkish EFL teachers' classroom assessment practices showed that most teachers relied on conventional assessment methods such as filling the blank, multiple-choice, matching,

and short-answer exams. Other formative methods showing a link between assessment and instruction, including group work, project, portfolio, performance assessment, and oral presentation, were used less frequently. In another example, Cheng and Sun (2015) investigated secondary school EFL teachers' classroom assessment practices in China, using a questionnaire with 350 junior and senior secondary school teachers. The findings indicated that teachers frequently used paper and pencil tests (e.g., major examinations and quizzes), and that performance and project-based assessments (e.g., projects by teams, projects by individuals, performance assessment) and teacher self-development assessments (e.g., oral presentations, objective assessments) were relatively less frequently used.

Teachers' practices of providing feedback

Research has shown that L2 teachers' feedback generally serves a summative function (Brumen et al., 2009; Gu, 2014; Lee, 2007b; Lee, Mak, & Burns, 2016; Saito & Inoi, 2017; Zhou & Deneen, 2016). For instance, Lee's (2007b) study revealed that secondary school EFL teachers provided feedback to inform students of their errors in writing. The feedback, in the form of grades and scores with no reference to specific criteria related to the goals of writing, was unlikely to help students obtain a clear understanding of the gap between their writing performance and the desired learning expectations. Brumen et al. (2009) also found that teachers provided numerical grades most frequently, while descriptive grades or comments were provided less often. They argued that, while giving numerical grades was fast, they provided no specific information on the quality of students' work. Saito and Inoi's (2017) study on Japanese junior and senior high school teachers' practices of classroom assessment reported that grading was the biggest concern for teachers when giving feedback. In contrast, there was little evidence of students' self-reflective feedback, peer feedback, and feedback on

discourse. The authors claimed that the impact on students was detrimental, as grades were not helpful in directing students' action for learning improvement, and also impacted negatively on students' confidence.

While most studies have demonstrated that teachers' feedback has not been fully used to support students' learning, Chen et al. (2014) provided evidence of the effective use of feedback. Using a case study approach, they investigated the assessment practices of two tertiary EFL teachers' assessment practices in China. The findings suggested that teachers generally provided supportive and timely oral feedback to students. Regarding the focus of feedback, while one teacher focused on identifying areas for improvement, the other went beyond this and gave detailed suggestions on how to improve.

Students' involvement in the assessment process

Several studies have looked into students' role in assessment (Chen et al., 2014; Cheng & Wang, 2007; Lee, 2007b; Xu, 2016), which suggests that students are not actively involved in the assessment process. Cheng and Wang (2007), for example, found that Canadian, Hong Kong, and Chinese ESL/EFL teachers took responsibility for undertaking classroom assessment and that students were neither engaged in preparing assessment criteria nor assessing their own and peers' learning. Chen et al. (2014) and Xu (2016)'s studies, which were based in the Chinese tertiary classrooms, also suggested that students were not actively involved in self-and peer assessment. Chen et al. (2014) indicated that while teachers had an awareness of the need to assign a role to their students in assessment practices, they still retained the power to evaluate students. Xu's (2016) study found that peers were not commonly utilized as sources of assessment by university EFL teachers; even when peer assessment was used the standards of achievement were not clarified, and students were asked merely to judge whether their

peers had done a better job. This practice undermines the positive effect of peer assessment. Lee's (2007b) study reported that students did not play an active role in the feedback process. It was found that teachers directly provided corrections for the errors in students' writing and that the students' role in identifying errors, locating the areas for improvements, and in engaging in error corrections was limited.

In summary, the review of current empirical research provides valuable insights into L2 teachers' CBA implementation, suggesting that little genuine, effective classroom-based assessment is taking place. The next section discusses the major challenges in CBA implementation in L2 contexts.

2.2.6.2 Challenges L2 teachers face in implementing CBA

A number of current studies have suggested that L2 teachers encounter challenges during the implementation of CBA, which can be categorized into three types: teacher-related challenges, student-related challenges, school and system-related challenges.

Teacher-related challenges

A major challenge is that L2 teachers have limited knowledge of CBA. Recent research has shown that there is a lack of training in classroom-based assessment for L2 teachers (Lam, 2015; Xu & Liu, 2009). It has been found that teacher education programs focus mainly on psychometric conceptions, placing little emphasis on CBA issues. Due to the lack of professional training in CBA, L2 teachers have reported a relatively low level of understanding of CBA. It is not surprising, therefore, that they have difficulty in implementing CBA effectively. As reported by Inbar-Lourie and Donitsa-Schmidt (2009), EFL teachers in Israel lacked expertise in using a range of alternative assessment methods (e.g., performance tasks, self-assessment, and peer assessment), which hindered them from using alternative assessment methods

effectively. Likewise, Brumen et al. (2009) found that Croatian teachers' limited knowledge of CBA restricted their use of authentic assessment methods such as portfolios and questionnaires.

Given the influential power of assessment literacy, researchers have suggested that teacher education programs should be provided to equip L2 teachers with professional skills and knowledge about classroom-based assessment (Chen et al., 2014; Inbar-Lourie & Donitsa-Schmidt, 2009; Lam, 2015; Lee & Coniam, 2013). Lam (2015), for instance, has stated that assessment training programs should be organized to provide pre-service teachers with adequate knowledge and skills in conducting both large-scale tests and classroom-based assessments.

Student-related challenges

The second major challenge relates to students' attitudes toward assessment and learning. For example, Lee et al. (2016) found that students' unfavorable attitudes towards peer feedback hindered teachers' feedback innovation. Chen et al. (2014) also found that as students believed that assessment was the responsibility of teachers, it was difficult for them to accept peer assessment. Researchers have argued that, for classroom-based assessment implementation to be effective, the challenge regarding students' attitudes must be addressed. Lee et al. (2016), for instance, has noted that teachers should organize discussions to help raise students' awareness of the benefits and problematic issues regarding innovative practices. Chen et al. (2014) have asserted that teachers' explanation of the learning philosophy behind classroom assessment can help students change their attitudes.

School and system challenges

The third set of challenges relates to the school and the educational system. At the school level, the implementation of classroom-based assessment is influenced by a variety of factors, including school policy, teacher appraisal, curricular demands, and class size. Studies have shown that the school policy, which is incompatible with formative CBA principles, may inhibit the implementation of CBA (Lee, 2007b; Lee & Coniam, 2013). For example, Lee and Coniam's (2013) study of EFL teachers' assessment practices in Hong Kong found that the school's assessment policy placed emphasis on providing feedback on all language errors. As a consequence, teachers were busy responding to every error, thus having few opportunities to provide effective feedback to move students forward. The teacher appraisal system has also been reported as a factor influencing teachers' CBA practices. Lee et al.'s (2016) study indicated that the amount of teacher feedback was evaluated by school leaders, which led to teachers tending to provide feedback on all errors, thus undermining efforts to maximize students' achievement with more useful feedback. Another school level challenge is the requirements of the curriculum. As reported by Lee et al. (2016), teachers' obligation to finish the rigid syllabus prescribed by their schools constrained their integration of assessment into their daily instruction. Other studies have shown that it is difficult to implement effective CBA in large classes as it is challenging for teachers to provide descriptive feedback for every student (Cheng, Rogers, & Wang, 2008; Cheng & Sun, 2015).

At the educational system level, the implementation of CBA faces two major challenges. One is the examination-driven culture (Butler, 2009; Cheng et al., 2004, 2008; Öz, 2014). It is acknowledged that many L2 school contexts are characterized by a culture of standardized testing, where there is a focus on using assessment for reporting,

administrative and certificate purposes (Lee et al., 2019). As a result of this examination culture, teachers are likely to place emphasis on summative scores and teach to tests, which limits their use of CBA to promote learning (Carless, 2011). For instance, Öz's (2014) study found that teachers' reliance on conventional assessments was influenced by the high-stakes final exams. Lee and Coniam (2013) also stated that in a culture that focuses on summative testing, it is hard to focus students' attention on formative assessment.

Another challenge relates to the Confucian heritage culture, which is prevalent in China and other nations that have been heavily influenced by the Chinese culture of learning (Nguyen, Terlouw, & Pilot, 2006). In the Confucian tradition, a heavy emphasis is placed on respect for teachers, which leads to a shared belief that the teacher is the authoritative figure and has the primary responsibility for the process of teaching and learning. Students, on the other hand, are regarded as passive recipients (Leng, 2005). This belief has led to a hierarchical student-teacher relationship, which is inconsistent with the principles of CBA (Carless, 2011). Research has shown that such a hierarchical relationship is a major barrier to the implementation of CBA in Confucian heritage contexts (Chen, 2009; Chen et al., 2014). Chen et al. (2014), for instance, found that college EFL teachers in China still dominated the classroom assessment environment and seldom actively engaged students in self-and peer assessment. Such practices were found to be influenced by teachers' beliefs about teacher authority rooted in Confucian heritage culture.

The review of the above studies suggests that L2 teachers generally have difficulty in translating the principles of CBA into classroom practice. The implementation of CBA is complex, and it faces a number of challenges, which are related to teachers, students, and school and educational systems.

In summary, four major gaps have emerged in the literature. First, current research has examined teachers' assessment practices that pertain to specific CBA principles such as learning goal setting, feedback provision, or student involvement. Few studies, however, have focused on these principles comprehensively. As noted by Lee et al. (2019), classroom-based assessment is a unitary concept, and thus, it is important to regard CBA principles as interdependent rather than individual or sequential entities. Thus, it is evident that a more comprehensive investigation into the implementation of CBA is needed.

Second, current empirical research has focused on the implementation and challenges associated with CBA. Little is known, however, about how teachers conceptualize CBA and how teachers' conceptions influence the way they implement CBA. This is an important gap given that teachers' conceptions are recognized in the literature as a crucial factor that may influence any pedagogical innovation (Borg, 2015).

Third, whereas the majority of previous studies have been conducted in secondary schools and universities, there is little information about classroom-based assessment of young language learners. Classroom-based assessment has been advocated as having great potential in facilitating young language learners' learning (Butler, 2019; Hung, Samuelson, & Chen, 2016). This study was designed, therefore, to investigate the implementation of CBA of young language learners.

Finally, in terms of research methods, much of the research is quantitative in nature, with a few studies looking into teachers' CBA practices in naturalistic settings. Since the implementation of CBA is a complex task that is influenced by a variety of factors, there is a need to investigate the implementation of CBA in context and in-depth.

Taking account of these gaps, the current study investigated teachers' conceptions and practices regarding CBA of young language learners through a mixed-methods approach. This part of the literature review has mainly addressed issues of CBA relevant to the present study. The next two parts draw on research in the fields of teacher conceptions and assessment of young language learners.

2.3 Teacher conceptions

Recent years have witnessed a growing interest in teachers' conceptions in both mainstream and language education (Barnes, Fives, & Dacey, 2015; Basturkmen, 2012; Borg, 2019; Couper, 2017; Freeman, 2002; Wu, Zhang, & Wei, 2019). This part of the review first introduces the origin of research on teacher conceptions. It then discusses various issues related to teacher conceptions, including definitions, importance, the relationship between teacher conceptions and classroom practice, and factors that influence the development of teacher conceptions. A review of research on L2 teachers' conceptions of classroom-based assessment is also presented.

2.3.1 Origin of teacher conceptions research

Prior to the 1970s, research into classroom teaching was dominated by a process-product approach (Borg, 2015). The assumption underlying this traditional paradigm was that students' learning was regarded as a product of teaching, and teaching was conceptualized as behaviors conducted by teachers (Borg, 2006). Researchers were, therefore, primarily concerned with the relationship between teachers' and learners' classroom behaviors (process) and students' learning outcomes (product) (Clark & Peterson, 1986). Little attention was paid to the role that teachers' cognitive processes played in influencing their classroom behaviors.

In the 1970s, there was a paradigm shift in research on classroom teaching, moving the focus from observable teacher behaviors to teachers' cognitive processes (Clark & Peterson, 1986). This paradigm shift was influenced by the development of cognitive psychology, which highlighted the impact of thinking on behavior (Calderhead, 1996). There was a growing recognition that teachers were not passive receivers of pedagogical knowledge but active decision-makers. This has, therefore, contributed to the domain of inquiry known as teacher conceptions (or teacher cognition). Research on teachers, consequently, focused not solely on teachers' behaviors, but also on understanding teachers' mental lives (Calderhead, 1996; Clark & Yinger, 1977; Fang, 1996).

2.3.2 Definitions of teacher conceptions

Although teacher conception has emerged as an independent research paradigm, formulating the terminology for teachers' cognitive processes remains difficult because of the number of terms used to refer to similar concepts (Borg, 2006, 2019; Pajares, 1992). For example, Borg (Borg, 2006) cites over 30 terms used in the literature to describe this cognitive process. Frequently used terms include *cognition* (Borg, 2003), *beliefs* (Watson, 2015), *knowledge* (Freeman, 2002), *perceptions* (Woods, 1996), *values* (James & Pedder, 2006), and *conceptions* (Brown, 2004).

Much of the debate on the terminology has surrounded the distinction between knowledge and beliefs. Some researchers have posited that *knowledge* is epistemologically distinct from beliefs because it is more objective and is associated with factual propositions, whereas *beliefs* is more subjective and is related to personal values (Barnard & Burns, 2012; Fenstermacher, 1994; Pajares, 1992). Fenstermacher (1994), for instance, has argued that "a claim to know something is epistemologically

different from simply having a belief in something” (p. 29). According to this perspective, knowledge can be understood as “a form of justified true belief” (Fenstermacher, 1994, p. 24); however, beliefs are closely concerned with “assumptions about the existence of entities” (Nespor, 1987, p. 321).

Some researchers have argued that there may be no clear-cut distinction between knowledge and beliefs (Andrews, 2003; Kagan, 1990; Verloop, Van Driel, & Meijer, 2001). They believe that both knowledge and beliefs can be seen as personal constructs that are concerned with how teachers themselves construct and understand concepts. From a constructivist perspective, Thompson (1992) has claimed that there is no priority of knowledge over beliefs, and vice versa. Her strong position is that it is more helpful to understand how these personal concepts impact teachers’ actions. She further proposed the concept of conceptions, allowing for the synthesis of beliefs and knowledge into one single construct to relieve the challenges researchers have faced in distinguishing the two concepts. According to Thompson (1992), teacher conceptions refer to “mental structures encompassing both beliefs and any aspect of the teachers’ knowledge that bears on their experiences, such as meanings, concepts, propositions, rules, mental images, and the like—instead of just beliefs” (p. 141). In this study, no distinction is made between knowledge and beliefs, and the term *conceptions* (Thompson, 1992) is adopted.

2.3.3 Importance of teacher conceptions

Research has shown that teachers’ conceptions are of considerable importance for two reasons. First, teachers’ conceptions seem to exert a powerful influence on teachers’ instructional decisions (Borg, 2001; Fang, 1996; Farrell & Ives, 2015; Kahn, 2000). Research has suggested that teachers hold conceptions about various aspects of teaching

and learning processes, such as conceptions about teaching, themselves, learners and learning, as well as subject (Basturkmen, 2012; Calderhead, 1996; Fang, 1996). These conceptions largely influence their instructional decisions, such as planning lessons, selecting teaching materials, designing instructional activities, and assessing students' performance (Burns, Freeman, & Edwards, 2015; Li, 2020).

Second, teachers' conceptions are deep-rooted and hard to change (Borg, 2003; Kagan, 1992; Li, 2017; Meschede, Fiebranz, Möller, & Steffensky, 2017). For example, Meschede et al. (2017) found that teachers' conceptions are stable and change slowly. Due to the stable structure, teachers' conceptions may outweigh the influence of teacher education and have a profound impact on teachers' professional lives.

Research on teachers' conceptions, therefore, is essential for our understanding of classroom teaching. From this perspective, understanding the relationship between teachers' conceptions of CBA and their assessment practices is essential, and, thus, was the main focus of the current study.

2.3.4 Teacher conceptions and classroom practice

The relationship between teacher conceptions and classroom practice has become an important topic of research. It is generally accepted that teachers' stated conceptions play an important role in their pedagogical practice. As Pajares (1992) claimed, teachers' beliefs "play a critical role in defining behavior and organizing knowledge and information" (p. 325). Likewise, Johnson (1994) argued that teachers' beliefs seem to be "instrumental in shaping how teachers interpret what goes on in their classrooms and how they will react and respond to it" (p. 440). Basturkmen (2012), in a review of research on the relationship between language teachers' conceptions and practices, also supports the notion that teachers' conceptions drive their actions.

Some empirical studies have demonstrated that teachers' conceptions influence the way they implement pedagogical practices (Bao, 2019; Farrell & Ives, 2015; Johnson, 1992; Kahn, 2000; Watson, 2015). For instance, Johnson (1992) found that ESL teachers' literacy instruction consistently reflected their theoretical beliefs. Bao (2019) examined Chinese as a second language teachers' beliefs and actual practices about corrective feedback. The study revealed consistency between their beliefs and practices in terms of the frequency of feedback, the emphasis on teacher-generated feedback, and feedback strategy.

While there is strong evidence that teachers' conceptions influence teachers' pedagogical practices, their practices do not always parallel their stated conceptions. As noted by Pajares (1992), teachers' conceptions are not always "a reliable guide to reality" (p. 326). Many studies have found an inconsistency between teachers' stated conceptions and their practices (Basturkmen, Loewen, & Ellis, 2004; Fu & Nassaji, 2016; Fung & Chow, 2002; Johnson, 1992; Wu et al., 2019). Fu and Nassaji (2016), for example, reported inconsistency between teachers' stated beliefs about the frequency of corrective feedback and their actual practice. Fung and Chow (2002) also found that novice teachers held positive attitudes towards a child-centered teaching approach, whereas their observed classroom practices were more teacher-centered.

A considerable body of studies has documented possible reasons for the contradiction between teachers' stated conceptions and their practices. Some studies have reported that the complexities of workplace contexts can hinder teachers from aligning their pedagogical practices with their conceptions. Such contextual complexities include class size (Jones & Fong, 2007), teaching syllabus and time constraints (Mori, 2011), and authority's influence, norms among colleagues, as well as parents' expectations (Andrews, 2003). Other reasons have been related to teachers,

including their varying teaching experience (Basturkmen et al., 2004), and personal linguistic background (Mori, 2011). Tensions between teachers' conceptions and practices have also been attributed to students' needs (Gilliland, 2015), personality (Roothoof, 2014), proficiency, and learning motivation (Graden, 1996).

Overall, there is evidence of a highly complex and dialectic relationship between what teachers think and what they do in the classroom.

2.3.5 Sources of teacher conceptions

Research has also identified a variety of sources that influence the development of teachers' conceptions. Research on the factors that influence how teachers establish their conceptions is particularly crucial in the context of teacher education as teachers' conceptions are unlikely to be established by reading assigned articles alone (Junqueira & Kim, 2013). According to Borg's (2003) schematic model of teacher cognition, these sources of influence can be divided into four categories: teachers' prior learning experience, teacher education, teaching experience, and contextual factors.

2.3.5.1 Teachers' prior learning experience

Researchers have argued that teachers' prior learning experience is an important influence on the formation of their conceptions (Borg, 2006; Grossman, 1991; Lortie, 1975). Lortie (1975) used the term *apprenticeship observation* to refer to the influence of teachers' prior learning experience on teachers' beliefs and practices. Lortie (1975) pointed out that students learn by observing their own teachers from primary school to university; these experiences are crucial in shaping their own conceptions of teaching and learning. Grossman (1991) also emphasized that teachers' prior learning experiences as students "makes it difficult for prospective teachers to consider alternative visions of teaching and learning" (p. 345).

Studies have provided empirical evidence that teachers' learning experience influence their conceptions about teaching and learning (Johnson, 1994; Moodie, 2016; Numrich, 2006; Woods, 1996). Johnson (1994), for example, found that preservice ESL teachers had strong images of their own teachers, teaching materials, and instructional activities, all of which had a powerful impact on their beliefs about what should be promoted or avoided in their own teaching. Moodie's (2016) study also showed that Korean EFL teachers' negative English learning experiences resulted in their intentions to adopt teaching models that were different from those they had experienced as learners.

2.3.5.2 Teaching experience

Another important source of teachers' conceptions is their own classroom teaching experience. As Crookes and Arakaki (1999) claimed, "accumulated teaching experience was the most often cited source of teaching ideas" (p. 16). It is believed that teachers accumulate experience of what pedagogical ideas, sources, methods, and content work effectively in their own circumstances throughout their teaching careers. These unique classroom teaching experiences, in turn, significantly reinforce or alter their conceptions about teaching and learning.

Various studies have highlighted the role of teaching experience in influencing teachers' conceptions (Breen, Hird, Milton, Oliver, & Thwaite, 2001; Junqueira & Kim, 2013; Kang & Cheng, 2014; Rahimi & Zhang, 2015; Sato & Kleinsasser, 2004). For instance, Rahimi and Zhang (2015), in their study of 20 experienced and 20 novice non-native English speaking teachers' conceptions about corrective feedback, found that experienced teachers held flexible conceptions while novice teachers held rigid conceptions about types and timing of corrective feedback. It appeared that experienced teachers, through reflecting on their classroom practices, believed that there was no

single solution for error correction. This study suggests that as teachers gain more teaching experience they value the role of corrective feedback in students' language development more than the novice teachers. Similarly, Akyel's (1997) study found experienced ESL teachers considered a wider range of instructional strategies to cater to students' needs, while novice teachers were less flexible in their choice of instructional strategies and were concerned more about the flow of instructional communication.

2.3.5.3 Teacher education

A third important influence on teachers' conceptions is teacher education. While some studies have reported limited impact of teacher education (Junqueira & Kim, 2013; Peacock, 2001; Urmston, 2003), there is increasing evidence showing that teacher education can influence teachers' conceptions (Borg, 2011; Cabaroglu & Roberts, 2000; MacDonald, Badger, & White, 2001). For instance, Cabaroglu and Roberts' (2000) study of 20 student teachers attending a 36-week *Postgraduate Certificate in Education* course identified development in teachers' beliefs in all but one teachers. The development of teachers' beliefs was the outcome of early awareness of pre-existing beliefs and self-regulated learning opportunities. Borg (2011) examined the impact of an eight-week in-service teacher education program on the beliefs of six English language teachers. There was evidence of considerable, although variable, impact on teachers' beliefs. Three teachers reported an awareness of their beliefs and could articulate key beliefs underpinning their work as a result of the teacher education program, while the other three teachers reported that they were not aware of any impact. Nonetheless, the results point to the potential impact of teacher education on in-service teachers' beliefs.

2.3.5.4 Contextual factors

Various studies have drawn attention to the powerful role that contextual factors play in shaping teachers' conceptions. Zhang and Liu (2014), for instance, in a study of Chinese junior secondary school EFL teachers' conceptions about English language teaching, found that school type had a direct influence on teachers' conceptions. The authors suggested that teachers working in schools with an innovative culture, supportive school authorities, qualified teachers, and high achieving students were likely to hold more constructive beliefs about collaborative learning and teacher-student interaction. Adams and Hsu's (1998) study suggested that the grade level at which teachers taught influenced their attitudes towards classroom assessment techniques. Primary school teachers of higher grade level (Grade 3 and Grade 4) attached greater importance to objective assessment techniques than teachers of lower grade level (Grade 1 and Grade 2). It has also been reported that teachers' interactions with their colleagues in their communities of practice, and even family members and friends in their personal lives, can shape their conceptions about teaching and learning (Barnard & Burns, 2012; Crookes & Arakaki, 1999; Zeichner & Tabachnick, 1985).

Having discussed the complex interplay between teachers' conceptions and classroom practice and the sources of teachers' conceptions, the next section focuses on L2 teachers' conceptions of CBA specifically.

2.3.6 L2 teachers' conceptions of classroom-based assessment

Despite the growing interest in teacher conceptions, research into L2 teachers' conceptions about CBA is scant (Davison, 2004). Nevertheless, the limited research has identified both convergence and divergence between L2 teachers' conceptions and practices regarding classroom-based assessment. There is evidence of the powerful

effect that teachers' conceptions have on the way they implement assessment practices (Chan, 2008; Mui So & Hoi Lee, 2011; Rogers, Cheng, & Hu, 2007; Wang, 2017). Mui So and Hoi Lee (2011), for example, in a study of 19 ESL secondary school teachers' conceptions and practices of assessment for learning, found that the teachers perceived assessment as evaluation-oriented rather than learning-oriented. Because of these beliefs, they tended to make the success criteria transparent to students. Rogers et al. (2007) found that tertiary EFL teachers in three research contexts (Canada, Hong Kong, Beijing) held strong beliefs, as reflected in their self-reported practices, about the role that assessment played in instruction and learning improvement. Chan (2008), in a survey of 520 primary school EFL teachers in Taiwan, also found that the stronger beliefs that teachers held towards multiple assessment methods, the more frequently they used multiple assessment methods in their daily teaching.

Other studies, however, have demonstrated that teachers' conceptions of classroom-based assessment are not always consistent with their assessment practices (Büyükkarcı, 2014; Chan, 2008; Chen et al., 2014; Guadu & Boersma, 2018; Troudi, Coombe, & Al-hamly, 2009; Xu & Liu, 2009; Zhou & Deneen, 2016). For instance, Chen et al.'s (2014) study of two university EFL teachers in China found that the teachers expressed positive attitudes towards the engagement of students in the assessment process. In contrast, in their actual practice, they seldom engaged students in self-and peer assessment. Shohamy, Inbar-Lourie and Poehner's (2008) large-scale survey study examined 467 teachers' conceptions and practices in regard to the assessment of Advanced Language Proficiency (ALP). They found that the teachers stated that the ALP construct could only be assessed through multiple assessment methods such as portfolios, performance tasks, essays, and rubrics and that assessment should be embedded in instruction to serve a formative purpose. In practice, however,

the predominant assessments used by teachers were traditional forms of assessment such as quizzes and tests. Similarly, Büyükkarcı's (2014) study of EFL teachers' assessment beliefs and practices in the Turkish context showed that while teachers expressed positive beliefs about the formative function of classroom assessment and the basic elements of formative assessment such as feedback, sharing learning goals, self-and peer assessment, they did not use formative assessment practices effectively in their classes.

Current studies have also identified multiple factors that contribute to the tension between teachers' conceptions and practices regarding classroom assessment. These include class size, prescribed curriculum (Büyükkarcı, 2014), education bureaucracy (Shim, 2009), professional training (Shohamy et al., 2008), as well as teachers' personal beliefs about the role of teachers (Chen et al., 2014). Inbar-Lourie and Donitsa-Schmidt (2009) provided an overview of factors that may hinder teachers from putting their assessment beliefs into practice. These are categorized into four types: cultural factors (social and cultural context of schools), political factors (authority's influence), technology factors (time allotment, availability of resources, assessment expertise) and postmodern factors (the authenticity, reliability and validity issues of assessment).

To sum up, two significant shortcomings have been identified in the studies reviewed. First, the investigation of teachers' conceptions of CBA has received limited attention, in comparison with research into the implementation of CBA. Studies such as Xu and Liu (2009) and Chen et al. (2014) focused mainly on the enactment of classroom-based assessment, thus providing limited information about how teachers' assessment practices have reflected their conceptions. It is crucial, therefore, to gain more information about teachers' conceptions to find ways to implement classroom-based assessment effectively. Second, while the limited research has offered some insights into how teachers' conceptions of CBA influence their assessment practices,

they often disregard the sources of teachers' conceptions. In other words, there is a paucity of research on the influence that previous language learning experience, teacher education, teaching experience, or contextual factors may have on shaping their conception. As previously mentioned, such investigation is of great importance for teacher education purposes.

2.4 Assessment of young language learners

2.4.1 Definition and characteristics of YLLs

Young language learners (YLLs) refers to “those who are learning a foreign or second language and who are doing so during the first six or seven years of formal schooling” (McKay, 2006, p. 1). Defining the age of YLLs can be a complicated issue because YLLs are introduced to second or foreign programs at a range of school years in different parts of the world (Bailey, 2008). For example, in most European countries, YLLs include students from both primary schools and the lower levels of secondary schools (Hasselgreen, 2005). In the United States, YLLs encompasses preschool and elementary school learners aged three to eleven years old (Bailey, 2008). Based on a review of previous research, in the present study, the term YLLs refers to learners who are learning a foreign or second language at the primary school level, with ages ranging between five to twelve years old.

Researchers have argued that YLLs are generally different from older learners or adult learners due to their unique characteristics (Bailey & Osipova, 2016; Butler, 2019; Carless & Lam, 2014; McKay, 2006). First, young learners are undergoing cognitive, social, and emotional growth. For instance, their memory and attention span is shorter than that of adult learners. It is easy for them to drop out of a task if they find it difficult. Second, young language learners are still acquiring literacy knowledge while learning

their first language. In contrast, adult learners have mature literacy knowledge. Finally, young learners are vulnerable to adults' praise and criticism.

Because of these special characteristics, assessing young language learners is believed to be a highly expert field of endeavor, requiring special consideration of test format, assessment type, task content, as well as assessment administration and interpretation (Bailey, 2008; Butler & Zeng, 2014; Hughes, 2003; McKay, 2006). For example, task type and content need to correspond to young learners' cognitive processing ability. Assessment tasks should be familiar to children to create a psychologically safe environment and should not be beyond children's first language understanding. It is also important that children experience a sense of success and progression when being assessed.

2.4.2 Research on YLL assessment

YLL assessment, as an independent research field, has just emerged in the field of language assessment. During the 1960s and 1970s, when early foreign language programs were first introduced in primary schools in the United States, Canada, and Europe, YLL assessment was an almost neglected area in the assessment landscape (Stern & Weinrib, 1977). This was largely due to the fact that no special assessments were used for young learners at that time; the assessment formats used for YLLs were adapted from standardized tests for the first language (L1) learners (Alderson & Banerjee, 2001). Later in the 1980s, criticisms of the adaption of L1 tests were expressed, such as Cummins (1981). According to Cummins (1981), there was a significant difference in language development between L1 learners and those pupils who were learning a second or foreign language. As a result, the formats of L1 tests were considered inappropriate for young language learners.

In the 1990s, YLL assessment came into its own as a research field, with a focus on large-scale testing for young pupils. During this period, foreign language programs for young learners were introduced worldwide (Alderson & Banerjee, 2001), which resulted in a great demand for large-scale tests at both school and national levels. Examples of such tests are the National Test of Reading Comprehension in Israel and the National Assessment Program in Education in Dutch, and the Practical English Level Test for Elementary English in Korea. These tests are designed to measure the language proficiency of young learners. They are categorized as summative assessments of language attainment and are often associated with high-stakes for both individuals and the school systems (Bailey, 2008). Because of their great consequences, large-scale tests for YLLs have received increasing research attention in terms of test validity (Hasselgreen, 2000; Schrank, Fletcher, Todd, & Alvarado, 1996; Szpotowicz & Campfield, 2016), and test impact on young learners (Choi, 2008; Morris, Lo, Chik, & Chan, 2000; Shohamy, 1998).

Since the start of the new century, classroom-based assessment has been acknowledged as a desirable means for improving YLL teaching and learning (Bailey, 2008; Brumen et al., 2009; Gattullo, 2000; Zangl, 2000). This reflects the trend in both general education and language assessment towards the alternative form of assessment (Genesee & Upshur, 1996; Rea-Dickins, 2001). Researchers have suggested that formative CBA encourages young learners to be actively engaged in learning and provides them with opportunities to become self-regulated learners (Butler, 2019; Hasselgreen, 2012; Saville & Weir, 2018). As reviewed previously, little is known about how teachers understand CBA and how they implement CBA to foster a close link between teaching, learning, and assessment of young learners. Given the crucial role that teachers play in classroom-based assessment (Butler, 2016), there is a need for more

studies on teachers' conceptions and practices regarding CBA so as to maximize its potential for advancing YLL learning.

In summary, echoing the words of Hasselgreen (2012), YLL assessment is a newly emerging field that warrants further research in a range of aspects. This study thus extends the limited research in the area of YLL assessment by investigating teachers' conceptions and practices of classroom-based assessment of young EFL learners in the Chinese context.

2.5 Summary

There are three salient gaps that arise from a review of the literature. First, CBA is increasingly advocated as having the potential to enhance teaching and learning. Previous research in general education has generated insights into the key principles for effective CBA. So far, however, there is a lack of a comprehensive understanding of the implementation of CBA in L2 contexts.

Second, teacher conceptions play a significant role in influencing their instructional behaviors. In educational contexts, there is substantial evidence that teachers' conceptions and practices represent a complex interplay, which is influenced by a wide range of factors. In addition, research has shown that teachers' conceptions are influenced by four major sources. To date, very little is known about L2 teachers' conceptions of CBA and the sources of their conceptions.

Third, CBA has been widely advocated to improve YLLs' learning. There is a paucity of studies, however, on how teachers conceptualize and implement CBA in the young learner context. There is, therefore, a need to look into CBA of young language to maximize the potential of CBA to support the learning of young learners.

The gaps in the research are the motivation for this study, which has investigated teachers' conceptions and practices regarding CBA of young EFL learners in the Chinese context. Specifically, this study aimed (1) to explore Chinese primary school EFL teachers' conceptions of the purposes and processes of CBA; (2) to examine the influence of teacher attributes (teaching experience, and previous education in assessment) and work environments (school type and grade level) on teachers' conceptions; (3) to investigate teachers' CBA practices; and (4) to understand the relationship between teachers' CBA conceptions and practices. As such, this study contributes to our understanding of teachers' conceptions and practices related to CBA within the context of teaching and assessing young EFL learners as well as how CBA can be effectively implemented in L2 contexts.

CHAPTER THREE

METHODOLOGY

3.1 Chapter overview

This chapter describes the research methodology starting with the research questions, the philosophical worldview, and the mixed-methods research design adopted in this study. The remaining two sections are concerned with participants, instruments, data collection, and analysis in two research phases. The final section outlines the ethical considerations relevant to the study.

3.2 Research questions

Before introducing the research design, the research questions are restated. The overriding aim of this study was to investigate the relationship between EFL teachers' conceptions and practices regarding classroom-based assessment of YLLs. The following four research questions address this aim.

RQ1: What are EFL teachers' conceptions of classroom-based assessment?

RQ1.1 What are EFL teachers' conceptions of the purposes of classroom-based assessment?

RQ1.2 What are EFL teachers' conceptions of the processes of classroom-based assessment?

RQ2: To what extent do teacher attributes (teaching experience and previous education in assessment) and work environments (school type and grade level) influence their conceptions of classroom-based assessment?

RQ3: How do EFL teachers implement classroom-based assessment practices?

RQ4: What are the relationships between teachers' conceptions and practices regarding classroom-based assessment?

3.3 Philosophical worldview

The term *worldview* refers to “a basic set of beliefs that guide action” (Guba, 1990, p. 17). It represents a cluster of philosophical assumptions and orientations about the world, social reality, and behavior, which, in turn, determines the approach to be adopted in a study (Creswell, 2014). Other synonymous terms include *paradigm* (Denzin & Lincoln, 2017) and *research methodology* (Nunan & Bailey, 2009).

Pragmatism was chosen as the overarching methodology for the current study. Pragmatists view the reality of the world as both singular and multiple (Creswell & Plano Clark, 2017). They believe that there must be a theory that works to explain a research phenomenon and that it is also essential to investigate individual cases to examine the nature of the phenomenon in depth. Applying these assumptions to research, pragmatists draw liberally from both quantitative and qualitative assumptions and collect both quantitative and qualitative data to provide a better understanding of a research problem (Biesta, 2010; Creswell, 2014; Mason, 2006).

Because of the large number of young EFL learners in China, a precursor for this study was to understand teachers' conceptions and practices regarding classroom-based assessment in general. Furthermore, as the implementation of classroom-based assessment is a complex issue influenced by a number of factors, individual teachers' assessment practices were examined to gain an in-depth understanding of the research problem. Pragmatism was adopted as fundamental to the inquiry as it enabled a general perspective as well as in-depth of teachers' conceptions and practices regarding CBA of young language learners.

3.4 Mixed-methods research design

This study adopted a mixed methods research design that involves the collection and interpretation of both quantitative and qualitative data (Biesta, 2010). Specifically, an explanatory sequential mixed methods design was adopted. It is a typical two-phase study in which quantitative data is first collected and analyzed, and the findings further explained with qualitative data in a subsequent phase (Creswell & Plano Clark, 2017). The quantitative phase allows a large amount of data to be collected so that the findings can be generalized (Cohen, 1988; McCusker & Gunaydin, 2015). However, the exploration of a complex research issue is constrained as it is not possible to obtain responses in depth. A sequent qualitative phase helps explore the data in a detailed and in-depth way, thereby contributing to a more comprehensive understanding of the issue (Denzin & Lincoln, 2017; Dörnyei, 2007). Such a design helps to generalize the research findings and also provides an in-depth understanding of the research phenomenon.

Mixed methods research also has weaknesses according to Morgan (2014). One is that it is more time consuming and expensive as it involves more procedures and participants. Another weakness is that it may be difficult for the researcher to interpret conflicting quantitative and qualitative data.

As an explanatory sequential design, this study was divided into two phases: a survey study and a case study (see Figure 3.1).

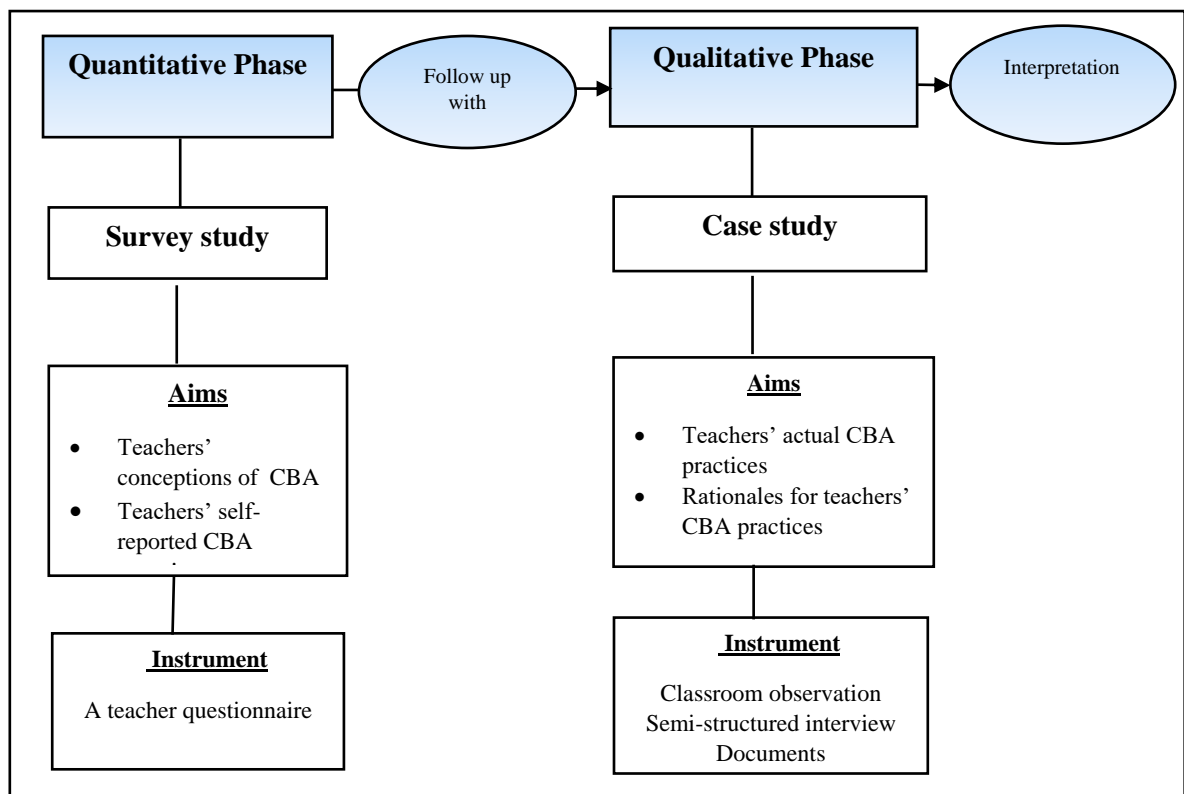


Figure 3.1 Overview of the research design

3.4.1 Phase One: Survey study

A survey study was adopted in Phase One to elicit EFL teachers' conceptions of classroom-based assessment and their self-reported assessment practices in general. Survey study provides a quantitative or numerical description of the characteristics of a population by studying a sample of that population (Fowler, 2013). Questionnaires, as a major type of survey study, have been widely employed to collect data about respondents' demographic characteristics, behaviors, and attitudes or opinions (Dörnyei, 2007). The merit of questionnaires lies in their capability to gather a large amount of information, based on which the findings can be generalized to a related population (Cohen, 1988). Thus, a questionnaire survey was carried out to enable the outcomes to be better generalized to a large number of primary school EFL teachers in China.

Although questionnaires can be efficient in collecting massive amounts of data, several limitations must be considered carefully. First, since questionnaire items are defined by the researcher, they may fail to cover the full range of the beliefs and knowledge held by the participants (Borg, 2006). Second, because participants usually need to respond to questionnaires without researchers' explanation, the participants may fail to interpret the items in a way intended by the researcher (Dörnyei & Taguchi, 2010). As a consequence, inaccurate responses may be provided. Participants, likewise, may not provide accurate responses because of the length of time it takes to complete questionnaires (Dörnyei & Taguchi, 2010).

Table 3.1 Research Design for Survey Study

Phase One — Survey study	
Major objectives	<ul style="list-style-type: none"> • To investigate teachers' conceptions of CBA purposes and processes • To investigate how teacher attributes and work environments influence teachers' conceptions of CBA • To investigate how teachers implement CBA practices • To examine the relationships between teachers' conceptions and practices regarding CBA
Participants	195 primary school EFL teachers
Data collection	An online questionnaire: <i>The Primary School English Teachers' Classroom-based Assessment: Conceptions and Practices Questionnaire</i> (See Appendix 1)
Data analysis	Exploratory factor analysis; descriptive analysis; independent samples t-test; one-way ANOVA; correlation analysis, multiple regression analysis

Specifically, the survey study was intended to investigate EFL teachers' conceptions of the purposes and processes of CBA, the effects of teacher attributes and

work environments on their conceptions, their CBA practices, and the relationships between teachers' CBA conceptions and practices. The *Primary School English Teachers' Classroom Assessment: Conceptions and Practices Questionnaire* was developed and administered to 195 primary school EFL teachers through convenience sampling. Table 3.1 presents an outline of the survey study design; a detailed description of the research methods is presented in Section 3.5.

3.4.2 Phase Two: Case study

A case study was employed in Phase Two to gain an in-depth understanding teachers' actual CBA practices and to elicit the rationales for their practices. Case study is defined as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between the phenomenon and the context are not clearly evident, and in which multiple sources of evidence are used" (Yin, 2003, p. 13). Three basic characteristics of case study are identified in the literature. The first is that a case is a bounded instance. That is, a case is regarded as a single entity with defined boundaries, either physical (such as a school, a child) or temporal (such as a lesson that has a beginning and an end) (Nunan & Bailey, 2009). Such an entity can be a single individual, a program, an institution, an organization, or a community. The second characteristic is that case study examines the characteristics of an individual entity in a naturally occurring situation, thus allowing for an in-depth investigation of the research problem (Simons, 2009). The particularity is another characteristic; that is, a case is selected precisely because the researcher intends to understand the particular, not to explore general characteristics of the many (Merriam, 1998).

A case study was appropriate, given the purpose of Phase Two. Although the survey study in Phase One identified general patterns of teachers' conceptions and

practices regarding CBA, it failed to provide a detailed and in-depth investigation of teachers' actual CBA practices and the rationales for their assessment practices. Since CBA is a complex issue, it was critical to look into teachers' CBA practices in the real world setting of the classroom. Therefore, the purpose of Phase Two was to learn about *how* teachers implemented assessment practices in a naturally occurring context and *why* teachers implemented such practices in the way they did. Case study methods are preferred when the research questions are focused on "how" and "why" issues (Yin, 2009).

The *definition of a case* is a fundamental issue in case study, which is conceptualized by Yin (2009) as *the unit of analysis*. In defining a case, it is important to identify an entity and specify the relevant information about the entity, since it is impossible to cover every aspect (Merriam, 1998; Nunan & Bailey, 2009). According to Yin (2009), the unit of analysis depends on the research questions. Therefore, this study regarded primary school EFL teacher's CBA practices in the Chinese context as the unit of analysis; that is, this study examined primary school teachers, not secondary school or tertiary-level teachers; EFL teachers, not ESL teachers; and teachers' CBA practices, not assessment practices that occurred outside the classroom.

Specifically, the case study was designed to provide insights into how EFL teachers actually implemented CBA practices and why they carried out the assessment practices the way they did. Three EFL teachers were selected through purposive sampling; they were invited to participate in classroom observations and semi-structured interviews. Relevant documents about classroom assessment were also collected. An overview of the design of the case study is presented in Table 3.2, and a detailed description of the research methods is provided in Section 3.6.

Table 3.2 Research Design for Case Study

Phase Two — Case study	
Major objectives	<ul style="list-style-type: none"> • To identify the characteristics of teachers' CBA practices • To gain insights into teachers' rationales for their CBA practices
Participants	3 primary school EFL teachers
Data collection	<ul style="list-style-type: none"> • Classroom observations with each participant • One semi-structured interview with each participant (after classroom observation data had been collected) • Relevant documents (e.g., teaching plans, the NECS, textbooks)
Data analysis	Thematic analysis

3.5 Phase One — Survey study

3.5.1 Participants

3.5.1.1 Sampling of participants

A definition of *population* and *sample* introduces this section. *Population* is the total number of participants that the study can draw from, while *sample* refers to the group of participants who participate in the research (Dörnyei, 2007). In the current study, given the large scale of English language education in primary schools, it was practical to study a smaller number of English teachers by using an appropriate sampling procedure.

A convenience sampling procedure was used in Phase One. Convenience sampling, as a non-probability sampling strategy, is most commonly used in L2 research (Dörnyei, 2007). By this sampling procedure, participants are recruited if they meet certain practical criteria, such as geographical proximity, availability at a given time, or

the willingness to participate in the study (Vehovar, Toepoel, & Steinmetz, 2016). The sampling procedure utilized in this study can be regarded as convenience sampling because the main objective was to collect information from primary school English teachers who could be approached easily. The target participants were primary school English teachers from two provinces in China, Chongqing and Guangdong. The researcher had access to the target participants as she had completed her undergraduate study in Chongqing during which she established a relationship with some primary school English language teaching advisors. Her research experience at the postgraduate stage also provided an opportunity to contact some teaching advisors in Guangdong. Therefore, it was convenient to approach participants in these two provinces.

As an obvious disadvantage of convenience sampling is the possibility of bias, researchers have suggested that convenience sampling should not be considered representative of the population (Etikan, Musa, & Alkassim, 2016). This problem was somewhat reduced in this study as participants were selected from two big provinces located in the Western (Chongqing) and Southern (Guangdong) parts of China, respectively. It allowed for the selection of participants across a broad geographical spectrum, thereby achieving greater generalizability.

3.5.1.2 Participants in survey study

Table 3.3 presents the demographic information for 195 participants: gender, age group, educational qualification, teaching experience, school type, grade level, and previous education in assessment. A majority of the participants were female teachers mostly under 40 years of age. The largest proportion of the participant teachers held a Bachelor's degree and a relatively large number had fewer than five years' experience in teaching English to young learners. Most teachers were from public schools and were

teaching Grade 3 and above. The majority of teachers reported that they had completed a course on assessment or had received training in assessment. Most of these teachers taught an average of four English classes, and were required to teach around 14 lessons every week.

Table 3.3 Demographic Information on Participants in Survey Study

Demographic variable	Groups	N	%
Gender	Male	6	3.1
	Female	189	96.9
Age	Under 30	51	26.2
	31-40	91	46.7
	41-50	45	23.1
	Over 50	8	4.0
Educational qualification	Under Bachelor	30	15.4
	Bachelor	153	78.5
	Master	12	6.1
Teaching experience	Less than 1 year	18	9.2
	1-5 years	48	24.6
	6-10 years	22	11.3
	11-15 years	45	23.1
	16-20 years	35	17.9
	Over 20 years	27	13.8
School type	Public	172	88.2
	Private	23	11.8
Grade level	Grade 1	14	6.7
	Grade 2	8	4.1
	Grade 3	49	25.1
	Grade 4	65	33.3
	Grade 5	67	34.4
	Grade 6	58	29.7
Previous education in assessment	No	69	35.4
	Yes	126	64.6

3.5.2 Development of questionnaire

The questionnaire used in Phase One comprises four sections: *Purpose Scale*, *Process Scale*, *Practice Scale*, and *Demographic Information* (see Table 3.4). The *Purpose Scale* included 20 items asking for teachers’ conceptions of CBA purposes. The *Process Scale* included 42 items asking for teachers’ conceptions of CBA processes, and the *Practice Scale* included the same 42 items requesting teachers to indicate how they implemented CBA. The fourth section asked teachers to provide demographic information (e.g., gender, age, educational qualification).

Table 3.4 Questionnaire Structure

Sections	Question items	Content	Question type
1 Purpose Scale	1-20	Teachers’ conceptions of CBA purposes	Likert scale
2 Process Scale	1-42	Teachers’ conceptions of CBA processes	Likert scale
3 Practice Scale	1-42	Teachers’ CBA practices	Likert scale
4 Demographic Information	1-9	Gender, age, educational background, teaching experience, school type, grade level taught, number of classes taught, English lessons taught per week, previous training or education experience in assessment	Multiple choice (1-6,9) open-ended (7-8)

Following Dörnyei and Taguchi’s (2010) guidelines, the questionnaire was developed in four steps: identifying salient constructs; generating item pool;

establishing rating scale; checking validity, translating and piloting the questionnaire. A detailed description of the procedures is presented in this section.

3.5.2.1 Identifying salient constructs

Questionnaire development starts with identifying critical constructs (Dörnyei & Taguchi, 2010). The *Purpose Scale* aimed to investigate teachers' conceptions of the purposes when assessing young learners. Informed by Brown (2006) and Brown, Kennedy, Fok, Chan, and Yu (2009), the *Purpose Scale* covered six potential constructs: (1) learning description; (2) learning improvement; (3) teaching improvement; (4) student accountability; (5) school accountability; (6) examination preparation. The first five potential constructs were identified with reference to Brown's (2006) model of conceptions of assessment. According to this model, four major assessment purposes are proposed: improvement, school accountability, student accountability, and irrelevance. The *improvement* purpose addresses the improvement of students' learning and the quality of teaching. It comprises four sub-constructs: assessment describes student learning, assessment improves student learning, assessment improves the quality of teaching, and assessment provides valid information. The sub-construct *assessment provides valid information* that was not included in the *Purpose scale* as it is concerned with the validity of assessment, such as whether assessment results were trustworthy and consistent. The *school accountability* purpose refers to the use of assessment results to demonstrate publicly that teachers or schools are doing a good job. The *student accountability* purpose concerns the use of assessment results to hold students individually accountable for their own learning. The *irrelevance* purpose addresses that assessment has a negative impact on students, and that assessment is used but ignored. This purpose was also excluded from this study because it is more closely linked to the effect of assessment on teaching and learning rather than the purpose of assessment. The

five potential constructs that were retained include learning description, learning improvement, teaching improvement, school accountability and student accountability. In addition, Brown et al. (2009) proposed that teachers use assessment practices for the purpose of preparing students for high-stakes examinations. As this purpose is relevant to the examination-driven culture of China, this was included as another potential construct *examination preparation* in the *Purpose Scale*.

The *Process Scale* explored how teachers conceptualized classroom-based assessment processes, and the *Practice Scale* was concerned with teachers' practices on the basis of these assessment processes. The potential constructs covered by the *Process Scale* and the *Practice Scale*, derived from Davison's (2008, as cited in Davison and Leung 2009) framework, were the same. The framework captures four key processes of teachers' classroom assessment: planning assessments, collecting information about students' learning, making professional judgments, and providing appropriate feedback. Based on McMillan's (2011) and Turner and Purpura's (2016) typologies of assessment methods, it was proposed that three types of assessment methods, spontaneous assessment opportunities, planned assessment opportunities, and formal assessment tasks could be employed to collect information about students' learning. Both descriptive feedback and evaluative feedback were included as it is important for teachers to distinguish the two major types of feedback (William, 2010a). Altogether, seven potential constructs are covered in both the *Process Scale* and the *Practice Scale*: (1) planning assessment; (2) using spontaneous assessment opportunities; (3) using planned assessment opportunities; (4) using formal assessment tasks; (5) making professional judgments; (6) providing descriptive feedback; and (7) providing evaluative feedback.

3.5.2.2 Generating item pool

Item generation refers to the translation of theoretically sound constructs into a set of items that explicitly measure a construct (Punch, 2009). A construct can be better captured by multiple items rather than a single item and a minimum number of three items need to be covered by each construct (Dörnyei & Csizér, 2012). To ensure that each set of items accurately reflects the latent construct, an item pool needs to be generated to include many more items than the final scale will contain. In the current study, the generation of the item pool was derived from two sources: items borrowed from established questionnaires and qualitative exploratory data with potential respondents (Dörnyei & Taguchi, 2010). First, some established questionnaires on classroom assessment were referred to as one source for the item pool, such as Brown's (2006) questionnaire on assessment conceptions and Cheng et al.'s (2004) questionnaire on classroom assessment practices. Second, semi-structured interviews were conducted with three primary school English teachers to elicit an in-depth understanding of their conceptions and practices regarding classroom-based assessment. An open, comprehensive interview topic is: What do you think about the purposes of teachers' classroom assessment and what assessment practices do you employ when assessing your students? Their responses were used as another source for the item pool.

For the *Purpose Scale*, a total of 22 items were generated. Five items were derived from teacher interviews, and the other 17 items originated from Brown's (2006) questionnaire. Some expressions from Brown's (2006) questionnaire were revised to ensure that the focus of the scale was on the purposes of assessment. For instance, "assessment is a way to determine how much students have learned from teaching (Brown, 2006, p. 168) was revised as "teachers use assessment to determine how much students have learned from teaching". In another example, "assessment places students

into categories” was revised as “teachers use assessment results to place students into categories according to their performance (e.g., high proficiency level and low proficiency level students)”.

For the *Process Scale* and the *Practice Scale*, a total of 44 items were initially generated. Some established questionnaires (Cheng et al., 2004; James & Pedder, 2006; Rogers et al., 2007; Zhang & Burry-Stock, 2003) were used as a source for the item generation. Although these established instruments do not focus on the dimensions of CBA processes directly, some questionnaire items do address certain aspects of CBA processes. For instance, the item “choosing appropriate assessment methods for instructional decisions” in Zhang and Burry-Stock’s (2003, p. 338) questionnaire emphasizes the selection of appropriate methods, an important element of planning assessment. The modified item used in the *Process Scale* was “Teachers select appropriate assessment methods according to students’ needs when planning language activities”, and the item in the *Practice Scale* was “I select appropriate assessment methods according to students’ needs when planning language activities”. Another source was the semi-structured interviews with three primary school English teachers. For example, one teacher stated that “I frequently check whether students have mastered what they learned in class through classroom tests, dictation and recitation”. This quotation influenced the development of items such as “Teachers collect evidence of learning through classroom tests” and “Teachers collect evidence of learning through dictation”.

3.5.2.3 Establishing rating scale

The *Purpose Scale* used a 6-point Likert scale, ranging from 1 for “strongly disagree”, 2 for “disagree”, 3 for “somewhat agree”, 4 for “agree”, 5 for “agree very much” to 6

for “strongly agree”. The major reason for the adoption of a 6-point scale rather than a 5-point scale was that respondents can use the middle rating on a 5-point scale to avoid definitive responses (Dörnyei & Taguchi, 2010). The *Process Scale* also had a 6-point Likert scale, ranging from 1 for “not important at all”, 2 for “not important”, 3 for “somewhat important”, 4 for “important”, 5 for “very important” to 6 for “completely important”. For the *Practice Scale*, teachers were asked to mark their responses to the same statements in the *Process Scale*, but on a different rating scale to identify how frequently they carried out various assessment activities. Specifically, it used a 6-point Likert Scale, ranging from 1 for “never”, 2 for “very rarely”, 3 for “rarely”, 4 for “occasionally”, 5 for “frequently” to 6 for “always”.

3.5.2.4 Checking validity, translating and piloting questionnaire

The questionnaire was revised several times to ensure its content validity. According to Patel and Joseph (2016), content validity refers to the degree to which the questionnaire items fully measure the construct. Typically, reviewers are invited to examine all the items for readability, clarity and comprehensiveness and come to an agreement as to which items should be included in the final questionnaire. In the current study, two doctoral supervisors and a group of postgraduate students from the field of language assessment were invited to provide feedback on the content validity. They examined the clarity, content and comprehensiveness of the three scales. Based on the feedback, two items were dropped from the *Purpose Scale*, and two items from the *Process Scale* and the *Practice Scale* were dropped; some items were revised.

The questionnaire was originally developed in English, but because it could be cognitively challenging for primary school EFL teachers to respond to the items in English, a Chinese version was prepared. To ensure the equivalence of the two versions,

a Ph.D. student, proficient in English and Chinese, was asked to translate the Chinese version back into English. The two language versions were compared in terms of language and content. When disagreement occurred, the items would be discussed and revised to ensure the Chinese translation was accurate.

The questionnaire was then piloted with 26 primary school English teachers. Piloting is an integral part of questionnaire development, as it helps identify problematic items and provides valuable feedback about the clarity of the instructions and the administration of the questionnaire (Derrick, 2016; Kiss & Nikolov, 2005). The questionnaire was sent to two primary school EFL teachers from another two provinces other than Chongqing and Guangdong from which the participants of the survey study were recruited. These two teachers sent the questionnaire to their colleagues who volunteered to complete it. In total, 26 questionnaires were returned, and the two teachers participated in a post-questionnaire interview, sharing about their understandings of the questionnaire items.

Two major implications arose from the pilot study, which are summarized as follows:

- *Modification of questionnaire items.* The teachers provided feedback on some items that were unclear to them; the items were revised accordingly. For instance, the phrase “with reference to the school’s requirements” was added to the item “teachers use assessment results to show how well they are doing” to emphasize the purpose of school accountability. In terms of completing the section on demographic information, the teachers reported that the item about ‘school type’ was too complex as the five options (private school, public school, foreign language primary school, township school, and urban school) can overlap.

According to this feedback, only two school types were retained: public school and private school.

- *Modification of questionnaire structure.* The demographic items were originally placed at the beginning of the questionnaire. One teacher, however, reported that this could make teachers feel like they were taking an examination. They recommended that demographic information be designed as the last section of the questionnaire. Thus, the overall structure of the questionnaire was revised.

Based on the pilot study, the questionnaire was finalized and was ready for administration. Appendices 1 and 2 provided the final English and Chinese versions used in the main study.

3.5.3 Data collection

Ethical approval was first gained from the University of Auckland Human Participants Ethics Committee. Following approval, six primary school English language teaching advisors were contacted to gain their consent to approach English teachers (four from Chongqing, two from Guangdong). Teaching advisors in China have the responsibility of guiding teachers' daily teaching activities and are in charge of all teachers in their districts. With their help, it was easy to approach potential English teachers. The six advisors were provided with information sheets that described the purpose and the process of the project, and consent forms for teachers to indicate their willingness to be involved in the project.

Once the six teaching advisors had been approached, an online invitation was sent out to all primary school English teachers within each district through the advisors. To ensure that a wide range of primary school English teachers was included in the final sample, teachers from different schools (both public and private schools) and grade

levels were invited to participate in the study. Those who decided to participate in the study were asked to complete the questionnaire during the weekend break and submit it online automatically. It took respondents approximately 30 minutes on average to complete the questionnaire. A total of 312 questionnaires were received in the end.

3.5.4 Data analysis

The data analysis consisted of data preparation and major statistical analyses. To be specific, data preparation included the input of data, checking missing data, outliers, and normality. Major statistical analyses were applied to answer the four research questions.

3.5.4.1 Data preparation

The first step of data preparation was to establish valid questionnaires. Questionnaires with an obvious response set (i.e., almost the same answers for all question items) were identified as invalid and were removed from the dataset. Of the 312 questionnaires collected, 117 were identified invalid. The data of the remaining 195 questionnaires (62.5%) was entered into the computer program IBM SPSS statistics version 24.

Checking missing data and outliers

No missing data were observed as the questionnaire items were designed as required questions. Cases with standardized scores in excess of 3 were identified as potential outliers, scores that are different from the rest, and distort statistics (Kline, 2011); that is, $|z| > 3.00$ indicated an outlier. As well as z scores, graphical methods such as histograms, box plots, or normal probability plots were used to detect outliers. Transformations were then undertaken to improve the normality of distribution by changing the score(s) on the variable for outlying cases (Tabachnic & Fidell, 2013).

Outlying cases were assigned with raw scores on the distorted variable; the score was one unit larger or smaller than the next most extreme score in the distribution.

Checking normality distribution

After outlier transformation, all variables in the three scales were checked to see whether they were normally distributed. Values of skewness and kurtosis were used to determine if the data were normally distributed (Kline, 2011). According to Kline (2011), absolute values of skewness of over 3.0 and kurtosis from about 8.0 to over 20.0 are described as “extreme”. In this study, an acceptable value of skewness was set between -3 and +3, and kurtosis between -8 and +8.

3.3.5.2 Major statistical analyses

Four major statistical analyses were used in this study. First, exploratory factor analysis (EFA) was conducted to explore the underlying constructs of the three scales. Second, descriptive analysis was used to describe the general characteristics of teachers’ CBA conceptions and practices. Third, independent samples t-test and one-way ANOVA were conducted to explore the differences in teachers’ CBA conceptions in regard to teacher attributes and work environments. Fourth, correlation analysis and multiple regression analysis were carried out to explore the relationships between teachers’ conceptions and their assessment practices. The following table provides an overview of the four major statistical analyses.

Table 3.5 Major Statistical Analyses

Major statistical analyses	Aims
Exploratory factor analysis	To explore the underlying constructs of three scales (Purpose scale; Process Scale; Practice Scale)
Descriptive analysis	To gain an overview of teachers' conceptions of CBA To gain an overview of teachers' CBA practices
Independent samples t-test & One-way ANOVA	To examine the effect of teacher attributes and work environments on teachers' conceptions of CBA
Correlation analysis & Multiple regression analysis	To examine the relationships between teachers' conceptions and practices regarding CBA

Exploratory factor analysis

Exploratory factor analysis was applied in this study to examine the underlying constructs of the three scales. Factor analysis is a useful technique for understanding the structure of a set of variables “by explaining the maximum amount of common variance in a correlation matrix using the smallest number of explanatory constructs” (Field, 2009, p. 629). EFA proceeds in an exploratory mode to identify the relations between sets of observed variables (indicators) and latent constructs (factors) (Byrne, 2010). It is appropriate when the links between the observed and latent variables are uncertain or unknown (Byrne, 2010).

Following the guidelines for carrying out EFA (e.g., Field, 2009; Osborne & Banjanovic, 2016), assumptions were first checked by examining Kaiser-Meyer-Olkin’s (KMO) Test of Sampling Adequacy and Bartlett’s Test of Sphericity. The KMO value examines the degree of common variance among questionnaire items, and a minimum

value of .50 is required to indicate that factors can be extracted (Kaiser, 1974). Bartlett's Test of Sphericity should be significant ($p < .001$) in order to provide evidence that a factor matrix can be extracted. The data were subjected to an EFA using solutions with non-orthogonal rotation (Direct Oblimin, $\delta = 0$). The Kaiser criterion, also known as "the eigenvalues > 1.0 rule", was selected to determine the number of retained factors. The pattern matrix was examined to check whether the factor structure had achieved structure simplicity. A factor structure was considered satisfactory given that: (1) each variable has salient loading (loading above .30 only on one factor; (2) each factor receives salient loadings from at least two variables; (3) there are no outlying variables (variables that seemed to be unrelated to other variables) (Tabachnic & Fidell, 2013). The name of the individual factor was determined by the common feature among the variables. All analyses were conducted using the statistical software IBM SPSS statistics version 24.

Descriptive analysis

Descriptive analysis summarizes findings by describing general tendencies and the variability of the scores in a dataset (Lodico, Spaulding, & Voegtle, 2010). A series of descriptive statistics were used to summarize teachers' conceptions of classroom assessment purposes and processes and their assessment practices. These statistics included the mean (M), standard deviation (SD) and percentage. The analysis procedures were completed using the statistical software IBM SPSS statistics version 24.

Independent-samples t-test and One-way ANOVA

The second research question investigated whether teachers' conceptions of classroom assessment were affected by *teaching experience*, *previous education in assessment*, *school type* and *grade level*. Independent-samples t-test can be used to compare two

means on the same dependent variable (Field, 2009). This method was employed to examine whether there were significant differences in teachers' conceptions of CBA in regard to *previous education in assessment* (those who had such experience and those who had no experience), *school type* (public school and private school) and *grade level* (lower grade level and higher grade level). One-way ANOVA was used to establish the effect of teaching experience because the independent variable *teaching experience* included three subgroups (Field, 2009): teachers with little experience (less than 5 years), medium experience (16 to 20 years) and high experience (over 20 years). Both independent samples t-test and one-way ANOVA were done using the statistical software IBM SPSS statistics version 24.

Correlation analysis and multiple regression analysis

Correlation analyses were performed to examine if there was a direct relationship between teachers' CBA conceptions and practices. Correlation analysis is a popular technique used for examining the direction and strength of the linear relationship between two variables (Pallant, 2011). The Pearson product-moment correlation coefficients (r) were applied to determine the relationships between teachers' CBA conceptions and their CBA practices. According to Pallant (2011), the sign in front of the correlation coefficient value determines whether there is a positive or negative relationship, and the size of the absolute value of the coefficient determines the strength of the relationship. As suggested by Cohen (1988), $r = .10$ to $.29$ indicates a small correlation, $r = .30$ to $.49$ indicates medium correlation and $r = .50$ to 1.0 indicates large correlation.

Multiple regression analysis allows the prediction of a single dependent variable from a set of independent variables (Pallant, 2011). It has been used to test the predictive

power of a set of variables on a particular variable and to assess the unique contribution of each independent variable. In this study, a series of multiple regression analyses were used to explore how well teachers' conceptions of CBA purposes and CBA processes were able to predict their CBA practices, respectively.

3.6 Phase Two — Case study

3.6.1 Participants

Phase Two focused on exploring teachers' actual CBA practices and the rationales for their practices in detail. Of the targeted participants, the 195 teachers from the survey study, 30 participants (19 from Chongqing, 11 from Guangdong) indicated their willingness to participate in the case study by providing contact information as requested on the questionnaire. Based on the consideration of the research budget and convenience, the case study was conducted in Chongqing with three teacher participants.

The purposive sampling strategy was used to select three, information-rich teacher cases from the pool of 19 potential participants from Chongqing (Merriam, 1998). Specifically, maximum variation sampling was adopted to optimize the exploration of variation within the primary school EFL teacher cohort to establish if there were shared patterns across the sampled diversity (Dörnyei & Taguchi, 2010; Suri, 2011). In selecting three teachers from the 19 potential participants from Chongqing, the attributes of teachers' age, grade level taught, educational qualifications, teaching experience, and previous education in assessment were considered. They were contacted through email and provided with information about the time commitment and how they would be involved in the study. All three teachers replied with consent to participate in the study. Table 3.6 provides an overview of the three teachers' demographic information.

Table 3.6 Demographic Information on Three Case Teachers

	Amy	Doris	Kathy
Gender	Female	Female	Female
Age	24	29	40
Educational background	Bachelor	Master	Bachelor
Major	Primary Education (English)	English Language Pedagogy	English Language Pedagogy
Teaching experience	2 years	4 years	12 years
School type	Public school	Public school	Public school
Grade level taught	Grade 3, 4, 6	Grade 4	Grade 6
English classes taught	4	6	4
English lessons per week	12	18	12
Previous education in assessment	Undergraduate course	Postgraduate course & in-service training	No

All three were female teachers working in public schools. Doris and Kathy worked in the same school in B district (pseudonym) of Chongqing, while Amy worked in another school in A district (pseudonym) of Chongqing. A more detailed description for each teacher is presented below.

Amy was a female teacher in her twenties. She held a Bachelor's degree in Primary Education (English). As a young teacher, she only had two years' English teaching experience at primary schools. She was teaching four classes, covering Grade 3 (2 classes), Grade 4 (1class) and Grade 6 (1 class). Each week, she needed to teach 12

English lessons. Amy reported that she had taken a course on assessment during her undergraduate study. However, she believed that the course content was theory-driven, providing limited practical guidance for their teaching practices.

Doris was also a female teacher in her twenties. She held a Master's degree and had been teaching English to primary school students for four years. Although she taught only Grade 4, she needed to teach all the classes at this grade level (6 classes together); altogether, she taught 18 lessons each week. Doris stated that she had not received any training sessions that were particularly targeted at classroom assessment during her in-service teacher training. She reported that there were only one or two sessions about assessment embedded in the teacher development programs that she had completed. In her opinion, what the lecturer taught was mostly theory-driven, with little guidance on how to put assessment theories into classroom practice. Doris mentioned that she had completed a course on assessment during her postgraduate study, but felt, similarly, that the course content had not been useful as it focused on analyzing examination papers for senior secondary school students.

Kathy was a female teacher in her forties. She held a Bachelor's degree in English language education and had 12 years' teaching experience. She was teaching four classes at Grade 6, and there were 12 lessons each week. In contrast to Amy and Doris, Kathy had received no training nor attended any course on assessment.

3.6.2 Instruments

3.6.2.1 Classroom observation

Classroom observation was employed in Phase Two to find out how EFL teachers implemented CBA practices in the natural classroom setting. Classroom observation was considered appropriate because it has the potential to explore the phenomenon

under investigation in depth (Esterberg, 2002; Hopkins, 2014). That is, first-hand information of assessment practices that had become routine to the participants themselves and could contribute to an enhanced understanding of the research phenomenon could be obtained. An additional motivation to use classroom observation was that observations could triangulate the findings of the survey study (Flick, 2018; Merriam, 1998). That is, potential differences between what teachers reported on their assessment practices and what they actually did with respect to the assessment of young learners can be established. Classroom observation makes it possible to see teachers' assessment practices as they are happening (Nunan & Bailey, 2009).

Although classroom observation has the advantage of collecting live data, a potential disadvantage is that it may result in reactivity on the part of those being observed (Hopkins, 2014; Lynch, 1996). The participants may change their behaviors when knowing that they are being observed, resulting in atypical rather than typical behaviors being displayed. In the current study, two strategies were adopted to minimize the negative effect of observations on teachers' behaviors. First, the researcher established a positive relationship with the three teachers through visiting their classrooms prior to the observations, so that the teachers would be aware of her role as an observer (Lincoln & Guba, 1985). Second, the purpose of observations was explained to the teachers to minimize any pressure that might be caused by the observations.

This study adopted the role of the non-participant observer; that is, the researcher observes the classroom setting with minimal or no involvement, enabling maximum time and flexibility to decide what to observe (Blaxter, Hughes, & Tight, 2010; Nunan & Bailey, 2009). The choice of non-participant observation was because if the observer is actively involved in the classroom activities, what happens in a natural classroom

setting might alter. Non-participant observations can minimize the negative impact of the presence of the observer.

Audiotaping was used, in this study, to record classroom interactions instead of videotaping, even though the latter is considered more powerful by capturing visual data because videotaping can be more intrusive and consequently may result in distorting the natural data (Nunan & Bailey, 2009). The three teachers agreed to put a recorder on the teacher's desk so that the interactions between teachers and students could be captured. Field notes, such as the number of students in a class, and materials used by teachers and students, were taken to supplement the audio-recording data. Lincoln and Guba (1985) suggest that field notes allow the observer to record his or her own comments that may be overhead and acted upon by the observer.

3.6.2.2 Interview

Interviews were used in Phase Two to explore teachers' rationales for their CBA practices. Teachers were asked to talk about the assessment practices they implemented and explain why they implemented the practices the way they did. Interview is defined as "a process in which a researcher and participant engage in a conversation focused on questions related to a research study" (DeMarrais, 2004, p. 51). The main purpose of interview is to find out what is "in and on someone else's mind" (Patton, 2002, p. 341). It is a natural way of collecting information from participants, which can be used to yield in-depth data. Interview was used in the current study, therefore, to gain an in-depth understanding of teachers' explanations of their own assessment practices.

Semi-structured interviews, which are a set of open-ended questions specified in advance, were adopted in this study (Merriam, 2016). This type of interview has several advantages (Corbin & Strauss, 2008; DeMarrais, 2004; Kvale & Brinkmann, 2015).

First, it provides participants with the flexibility to follow their own flow of thoughts, which helps elicit implicit explanations for their practices. Second, the semi-structured nature of interview questions allows for probing to clarify understandings of participants' comments about their practices. Semi-structured interviews, however, are time-consuming in terms of both data collection and analysis (Alshenqeeti, 2014). Moreover, the flexible use of open-ended questions may affect the way different participants respond to the same questions, which can reduce comparability (Adams, 2015).

An interview guide was used in the current study to ensure that the research domain was properly covered. It was framed around two sections: (1) teachers' background information; (2) questions related to teachers' CBA practices, including the purposes for teachers' assessment practices, teachers' practices of planning assessment, collecting learning evidence, making professional judgments, and providing improvement feedback. Appendix 3 provides a list of questions covered in an interview.

3.6.2.3 Documents

Relevant documents were collected in Phase Two to supplement data on teachers' CBA practices in the naturalistic setting. Documents refer to "a wide range of written, visual, digital, and physical materials relevant to the study at hand" (Merriam, 2016, p. 139). Lincoln and Guba (1985) suggest that documents can provide a rich source of information that is contextually relevant and grounded in the contexts they represent. Document data were collected in this study to help understand the context in which teachers carried out their assessment practices and to help probe in-depth into teachers' actual assessment practices.

3.6.3 Data collection

Three teachers in Chongqing were recruited from those who had completed the questionnaire and were willing to participate in the case study purposively. Classroom observations were conducted with each teacher four times, spanning a period of six weeks. The researcher, as a non-participant observer, was sitting in the back of the classroom and taking no part in the classroom activity. All classroom observations were audio-recorded, with participants' permission, with a recorder on the teacher's desk to capture interactions between teachers and students. Field notes were taken to supplement the audio recordings. Each teacher was observed for four classes for a total of around 160 minutes. After each observed class, relevant documents were collected to help better understand teachers' assessment practices. Those documents included copies of teachers' lesson plans, textbooks, and students' exercise books, and teaching slides.

Table 3.7 provides detailed information about the classroom practice of the three teachers being observed. Amy was teaching three grade levels, Grade 3, Grade 4 and Grade 6, while Doris and Kathy were teaching Grade 4 and Grade 6, respectively. To enable a comparison of teachers' practices at different grade levels, the grade level observed for Amy was Grade 3. To be specific, Class 6 with 47 students, randomly selected by Amy, was observed four times. Amy used the *PEP English* (Grade 3 Book 2), a textbook published by the People's Education Press, which consisted of six learning units and two review units. Doris' Grade 4 Class 1, Class 1 and Class 1 were observed. There was an average number of 40 students in these classes. Doris' school used another textbook, the *Standard English* published by the Foreign Language Teaching and Research Press, which comprised ten learning modules and one review module. In Kathy's case, three classes of Grade 6 (Class 1, Class 3, and Class 5), with

an average number of 50 students in each class, were observed. The textbook used by Kathy was also the *Standard English*.

Table 3.7 Background Information on Classroom Observations

Teacher	Grade level	Class	N	Textbook	Lesson
Amy	Grade 3	Class 6	47	<i>PEP</i>	Unit 3 At the Zoo: let's spell
		Class 6	47	<i>English</i>	Unit 3 At the Zoo: What is it?
		Class 6	47	(Book 2)	Unit 4 Where is my car? Let's learn
		Class 6	47		Unit 5 Do you like apples? Let's learn
Doris	Grade 4	Class 1	40	<i>Standard</i>	M 3 Unit 2 Days of week
		Class 1	40	<i>English</i>	M 3 Unit 2 Days of week
		Class 5	42	(Book 2)	M 4 Unit 2 Will it be hot in Haikou?
		Class 2	43		M 4 Unit 2 Will it be hot in Haikou?
Kathy	Grade 6	Class 1	50	<i>Standard</i>	M 4 Unit 1 The apples are falling down the stairs
		Class 3	51	(Book 2)	M 8 Unit 1 Why do you have cups on your heads?
		Class 3	51		M 8 Unit 1 Why do you have cups on your heads?
		Class 5	52		M 8 Unit 1 Why do you have cups on your heads?

Note: N = Number of students in each class; M refers to a module in the textbook

Semi-structured interviews were conducted with each individual teacher within two days after the final observation of the class. The interview time was chosen by teachers at their convenience: two teachers were interviewed during lunch break and one teacher after school. One teacher was interviewed in the office and another two in the classroom. Interviews were conducted in Mandarin as it is generally suggested that interviews are better conducted in interviewees' mother language (McDonough &

McDonough, 1997). During the interview, the interview guide was used to ensure that the major questions were covered, while probes were appropriately used to ask for more details and for clarification. With the permission of teachers, interviews were audio-recorded to ensure that everything said was preserved for analysis. Notes were also taken during each interview to signal important information noted by the interviewees and to record reactions to something the teachers said. The three teachers were interviewed individually for approximately one hour.

3.6.4 Data analysis

Within qualitative inquiry, data analysis is the process of making sense of the data (Miles, Huberman, & Saldaña, 2014). It involves categorizing all the data into meaningful 'bits' of data, reconstructing the 'bits' of data to generate abstract themes, representing the themes, and making an interpretation of the themes. The process of qualitative data analysis is complex and iterative and involves moving back and forth between concrete units of data and abstract concepts (Merriam, 2016).

Specifically, a thematic analysis approach was used to identify themes or patterns within the data in this study. Thematic analysis is an analytical method for segmenting, categorizing, summarizing, and reporting patterns within a qualitative dataset (Braun & Clarke, 2006). Thematic analysis was appropriate for Phase Two because it helped group the rich data from the observations and interviews into meaningful units so that any patterns of teachers' CBA practices, and the rationales for their practices, could be identified.

Following Braun and Clark's (2006) guidelines for conducting thematic analysis, the analysis of qualitative data in this study was composed of three major steps: processing data, generating initial codes, and generating themes and going beyond.

3.6.4.1 Processing data

The first step was to process the qualitative data. As qualitative data analysis focuses on words as the basic form, the raw data such as recordings needs to be processed first (Miles et al., 2014). Although the transcription process is time-consuming, it helps develop a thorough understanding of the data (Lapadat & Lindsay, 1999).

In the current study, twelve audio recordings of observations by three teachers were fully transcribed into text. As teachers' instruction during one class usually consisted of a series of episodes, each transcript was divided into different episodes according to the instruction focus. This made it easier to generate codes from a mass of data. For instance, one episode at the beginning of a class was named "assessment: oral reading word cards" because the focus of the instruction was on checking students' attainment through oral reading. Appendix 4 provides an example of the coding procedures of observation data. During the transcription process, the recordings were compared with the field notes, and memos were written by the researcher (Miles et al., 2014). Repeated reading of the transcripts was made afterward to enable the researcher to become familiar with all aspects of the data (Braun & Clarke, 2006).

All three audio recordings of the interviews were entirely transcribed into texts. While listening to the recordings, notes of initial thoughts on the data were recorded. After each recording was transcribed, the transcript was checked against the original recording for any discrepancies. Likewise, the transcripts were read and re-read to get an overall picture of the data. The transcripts were analyzed in Chinese, the language in which they were conducted, and only the codes and themes were translated into English. This was because the difference between the source language and the target language

can cause loss of information in the process of translation, which may affect the accuracy of the analysis.

3.6.4.2 Generating initial codes

Generating codes constitutes a fundamental step in analyzing qualitative data. The process of coding involves segmenting the data into chunks and labeling those chunks with a term (Creswell, 2014). Boyatzis (1998) describes a code as “the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon” (p. 63).

A combination of deductive and inductive coding approach was used in the current study (Miles et al., 2014). A list of prior coding categories was developed, drawing on the literature review and the findings of the survey study. The coding categories for all observation data, together with documentary evidence, specifically focused on the main components of CBA practices, which included planning assessment, using spontaneous assessment opportunities, using planned assessment opportunities, using formal assessment tasks, making professional judgments, and providing feedback. The coding categories for interview data focused on major factors that affected teachers’ CBA practices, including teacher-related factors (e.g., teachers’ conceptions, teaching experience, previous training experience), student-related factors (e.g., students’ proficiency) and contextual factors (e.g., class size, exam pressure). An inductive approach was applied, allowing for new ideas to emerge progressively during the process of coding.

3.4.6.3 Generating themes and going beyond

The next step was to generate themes, which involves sorting a list of different codes into a small number of potential categories (Braun & Clarke, 2006). Essentially, this is

a process of looking for an overarching pattern in the data (Miles et al., 2014). It was a recursive and interactive process involving multiple interpretations of the data, where the research question had provided a domain of relevance. In this way, the recurring patterns of teachers' CBA practices and the rationales for their assessment practices could be identified. Figure 3.2 presents an example of how the overall pattern of teachers' practices of planning assessment was generated.

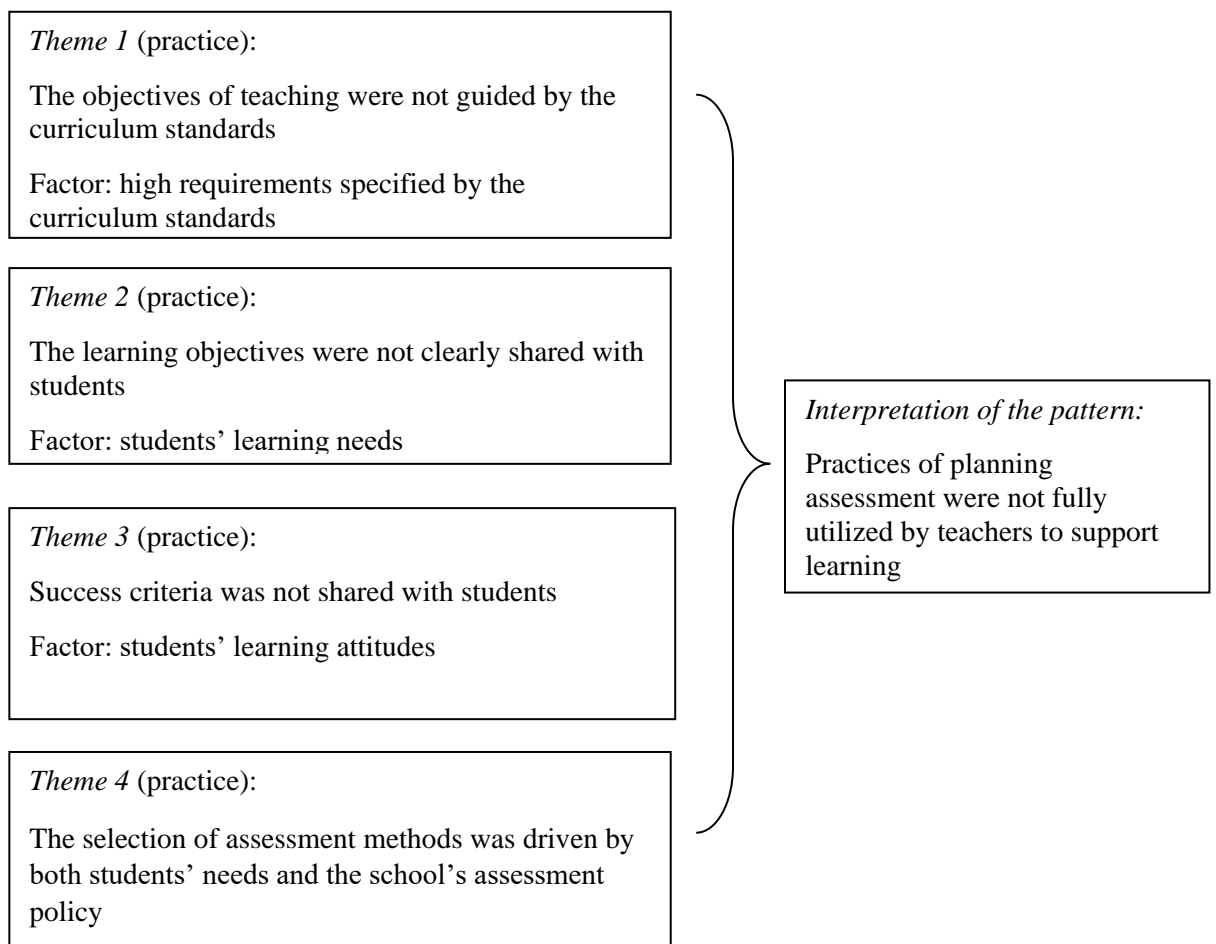


Figure 3.2 Example of generating the pattern of planning assessment

Inter-rater reliability was evaluated by analyzing a sample of transcribed observation and interview data (Belotto, 2018). A sample of codes and themes was selected and given to a Ph.D. student whose research interest was classroom assessment and who was also using qualitative research techniques for analysis. She agreed on most

of the categorization, with an inter-coder agreement of 87.0%. The remaining differences were negotiated to agree on a final decision.

3.7 Ethical considerations

This study had been approved by the University of Auckland Human Participants' Ethics Committee. Major ethics issues that might arise from this study included voluntary participation, informed consent, participants' rights to withdraw, and anonymity and confidentiality. This section gives a brief account of these ethical considerations.

3.7.1 Voluntary participation

Participants were informed that participation in the project was voluntary, and they could withdraw from the project at any time during the data collection. They were provided with a Participating Information Sheet and a Consent Form, outlining clearly what they were expected to do in the study. During the data collection, participants had the right to refuse to complete the questionnaire, to be observed, or to answer interview questions. Before carrying out classroom observations and semi-structured interviews, the participants were informed that the process would be audio-recorded. They had the right to refuse to be recorded, and they could ask to have the device turned off at any time. Compensation was made to the participants for their time and effort.

3.7.2 Informed consent

Participants were provided with adequate information about the research, including the purpose, methods of participant involvement, and intended use of the results, to ensure that they were only participating in the study given their informed consent. The

information was provided in a manner that could be easily and effectively understood by the participants.

3.7.3 Participants' rights to withdraw

Participants were clearly told they had the right to withdraw from the research project at any time during the data collection period, without giving a reason. It was made clear to them that, once questionnaire data had been collected, data could not be withdrawn because questionnaires were anonymous. Participants were informed they were entitled to withdraw the classroom observation and interview data they had provided within a period of two weeks after data collection. After that time, it was not possible to withdraw the data because the analysis of the research results was underway.

3.7.4 Anonymity and confidentiality

Anonymity was guaranteed for the online questionnaire; to better preserve anonymity, participants were not required to write their names on the questionnaires. For classroom observations and interviews, pseudonyms were used when reporting specific findings in order to protect the identity of each participant. To protect the confidentiality, the data collected would be accessible only to the researcher and her supervisors.

3.8 Chapter summary

This chapter has presented an overview of the methodology, including the research questions, philosophical worldview, and research design. A detailed description of participants, instrument, data collection, and analysis of the two phases has been provided, respectively; major ethical considerations related to this study have also been addressed. In the following two chapters, results regarding the survey study and the case study are presented respectively.

CHAPTER FOUR

RESULTS OF THE SURVEY STUDY

4.1 Chapter overview

This chapter reports the results of the survey study, which are presented in relation to the four research questions. The underlying constructs of the *Purpose Scale*, the *Process Scale*, and the *Practice Scale* are presented first. The ensuing four sections report teachers' self-reported conceptions of classroom-based assessment, the effect of teacher attributes and work environments on their conceptions, their classroom-based assessment practices, and the relationships between their conceptions and practices elicited in the questionnaire.

4.2 Exploring underlying constructs: CBA purposes, processes and practices

The questionnaire consisted of three scales that were developed to elicit teachers' conceptions of CBA purposes, their conceptions of CBA processes, and their CBA practices, respectively. This section describes the underlying constructs of each scale, which were explored through EFA.

4.2.1 Classroom-based assessment purposes

Descriptive statistics for the *Purpose Scale* showed that the mean scores of the 20 items ranged from the lowest, 3.15 (Q1.9), to the highest, 5.19 (Q1.7), with standard deviations ranging from .634 to 1.344. The values of skewness and kurtosis were within the cut-off values of |3.0| and |8.0| respectively (see Appendix 5), indicating the univariate normality of the responses (Kline, 2011). The total sample was then subjected to exploratory factor analysis.

Table 4.1 EFA Results of Purpose Scale

Factors	Items	Factor loading			α
		F1	F2	F3	
Learning improvement	Q1.2	.916			.854
	Q1.1	.804			
	Q1.3	.734			
	Q1.5	.649			
	Q1.7	.577			
	Q1.6	.575			
Student accountability & examination	Q1.9		.872		.812
	Q1.16		.763		
	Q1.8		.706		
	Q1.15		.699		
	Q1.10		.654	.	
Teaching improvement	Q1.13			.887	.873
	Q1.12			.837	
	Q1.14			.833	
	Q1.11			.748	

In the process of EFA, two items (Q1.4 and Q1.18) achieved a factor loading of over .30 on more than one factor and were thus discarded; one item (Q1.17) was identified as an outlying variable; one factor comprised only two items (Q1.19 and Q1.20) and was thus deleted. In the final data, three factors were extracted from the retained 15 items ($KMO = .836$, $df = 105$, $p < .001$). A significant test result for Bartlett's Test of Sphericity ($p < .001$) provided evidence that correlations between items were sufficiently large for the extraction of a factor matrix. The KMO value of .836 suggested that the items shared a very high degree of common variance. The total variance explained by the three factors was 63.54%, which was considered satisfactory.

Further, Cronbach's alpha coefficient for three factors ranged from .812 to .873, indicating good internal consistency. Based on the commonalities shared by the items, the three factors were named as follows: Factor 1 *Learning improvement* (37.57 % variance); Factor 2 *Student accountability and examination* (17.97% variance); Factor 3 *Teaching improvement* (8.00% variance). Table 4.1 presents the three factors, the component items, factor loading, and alpha coefficients.

The *Purpose Scale* was initially constructed to cover six latent factors, including *learning description, learning improvement, teaching improvement, student accountability, school accountability, and examination preparation*. The results of EFA identified a three-factor structure with 15 items. Initial items about learning improvement and learning description were loaded onto one factor. This factor was then interpreted as learning improvement because all items were related to the use of assessment to help describe and enhance students' learning. Specifically, four items concerned the purpose of describing and diagnosing student learning, such as using assessments to determine what they had learned and to identify strengths and weaknesses in their learning. Two items were related to the purpose of using assessment to enhance students' learning motivation.

Initial items about student accountability and examination preparation were also loaded onto one factor. This factor was interpreted as student accountability and examination. It reflected the purpose of making students accountable, such as placing them into categories, assigning a grade or level to students' work, and using assessment results to determine if students had met certain standards. It also included the purpose of preparing students for examinations using assessment tasks. For instance, teachers designed language activities according to the requirements of public examinations and taught students examination-taking skills to help them pass exams.

The third factor was named teaching improvement purpose. The main concern of the four items loaded on this factor was to help teachers improve their daily instruction, such as identifying strengths and weaknesses in teaching and modifying the instruction.

4.2.2 Classroom-based assessment processes

Descriptive statistics for the *Process Scale* showed that the mean scores of 42 items ranged from the lowest, 4.58 (Q2.17), to the highest, 5.37 (Q2.1), with standard deviations ranging from .589 to .778. The absolute values of skewness and kurtosis were within the cut-off values of |3.0| and |8.0| respectively (see Appendix 6), which indicated the univariate normality of the responses. Accordingly, the sample of 195 responses was subjected to EFA.

During the process of factoring, five items were loading on more than one factor (Q2.31, Q2.12, Q2.7, Q2.28, and Q2.26) and were thus dropped from the factor analysis. In the final solution, seven factors were successfully extracted, accounting for 72.54% of the total variance ($KMO = .934$, $df = 666$, $p < .001$). The Cronbach alpha coefficient for the seven factors ranged from .834 to .953, indicating very good internal consistency among items within each factor. It also indicated that the factor structure could be used as a reliable scale for further analysis. Based on the shared commonalities, the seven factors were named: F1 *providing descriptive feedback* (45.39 % variance); F2 *using student-involving assessment opportunities* (7.60 % variance); F3 *using formal assessment tasks* (5.10% variance); F4 *providing evaluative feedback* (4.33% variance); F5 *planning assessment* (4.30% variance); F6 *making professional judgments* (3.38% variance) ; F7 *using spontaneous assessment opportunities* (2.72% variance). The factor names, items, factor loading, and alpha coefficients are presented in Table 4.2.

Table 4.2 EFA Results of Process Scale

Factors	Items	Factor loading							α
		F1	F2	F3	F4	F5	F6	F7	
Providing descriptive feedback	Q2.42	.890							.953
	Q2.41	.880							
	Q2.40	.854							
	Q2.39	.828							
	Q2.38	.742							
	Q2.37	.696							
Using student-involving assessment opportunities	Q2.27	.318							.911
	Q2.19		.807						
	Q2.18		.799						
	Q2.17		.653						
	Q2.20		.638						
Using formal assessment tasks	Q2.16		.546						.886
	Q2.15		.540						
	Q2.22			.809					
	Q2.23			.766					
	Q2.24			.718					
Providing evaluative feedback	Q2.21			.641					.834
	Q2.25			.352					
	Q2.34				.754				
	Q2.33				.602				
	Q2.35				.568				
Planning assessment	Q2.36				.394				.847
	Q2.1					.788			
	Q2.2					.721			
	Q2.5					.519			
	Q2.3					.427			
	Q2.4					.394			
Making professional judgments	Q2.6					.305			.889
	Q2.30						.809		
	Q2.29						.712		
Using spontaneous assessment opportunities	Q2.32						.603		.881
	Q2.11							-.707	
	Q2.10							-.694	
	Q2.8							-.647	
	Q2.9							-.592	
Using assessment opportunities	Q2_14							-.435	.881
	Q2_13							-.371	

The *Process Scale* was developed to measure teachers' conceptions of the processes of classroom-based assessment. It was initially designed to cover seven

potential constructs: planning assessment, using spontaneous assessment opportunities, using planned assessment opportunities, using formal assessment tasks, making professional judgments, providing descriptive feedback, and providing evaluative feedback. The results of EFA resulted in a seven-factor solution, which, to a large extent, confirmed the initial constructs.

The first factor was labeled providing descriptive feedback, which included seven items. This factor was concerned with providing feedback to help students improve learning, such as providing feedback to help identify strengths in learning, identifying students' weakness in regard to the learning goals, helping students find ways of solving problems, and helping students understand what they need to do to improve their work.

The second factor included six items regarding collecting evidence through student-involving assessment opportunities. These assessment opportunities were usually embedded in regular instructional activities; it was less apparent to students that these activities were used as assessments. Such assessment opportunities placed emphasis on students' performance, or students' role, in the assessment process, and included self-assessment, peer assessment, portfolios, oral presentations, and other language activities.

The third factor, in contrast, concerned formal assessment tasks. It reflected the use of more formal types of assessment, which were often planned in advance and were more apparent to students. These assessment methods included classroom tests, dictation, assignments, and oral reading and reciting.

The fourth factor was about providing evaluative feedback on students' learning achievement. The main purpose was to provide feedback to students and their parents about what they had achieved, through scores, grades, written, or oral comments.

The fifth factor, with six items, was named planning assessment. Four items were related to instructional objectives, such as establishing instructional objectives, designing instructional objectives according to curriculum requirements and students' needs, and helping students understand the objectives. Another two items were about establishing and sharing assessment criteria with students.

The sixth factor, with three items, was named making professional judgments as it reflected how the judgment about students' performance was made in a valid and reliable way. Two items were about comparing students' performance against pre-set learning goals and their previous performance, while one item was about looking for overall patterns of students' learning while interpreting assessment information.

The final factor, with six items, was labeled using spontaneous assessment opportunities. These assessment opportunities were informal, unplanned, and incidentally embedded in teachers' daily classroom instruction and were mainly used to modify instruction and offer immediate feedback to students. Such assessment opportunities included teacher questioning and informal conferences with students, teachers, and parents after class.

4.2.3 Classroom-based assessment practices

Descriptive statistics for the *Practice Scale* showed that the mean scores of the 42 items ranged from the lowest, 4.33 (Q3.15), to the highest, 5.46 (Q3.1), with standard deviations ranging from .579 to .994. The values of skewness ranged from -.732 to -.015, and the values of kurtosis ranged from -.634 to 1.237. Both values were far from the cut-off values of ± 3.0 and ± 8.0 , respectively (see Appendix 7), indicating the univariate normality of the responses. The responses were thus subjected to EFA.

Table 4.3 EFA Results of Practice Scale

Factors	Items	Factor loading						α
		F1	F2	F3	F4	F5	F6	
Providing descriptive feedback	Q3.41	.952						.954
	Q3.38	.865						
	Q3.42	.845						
	Q3.40	.803						
	Q3.39	.754						
	Q3.37	.722						
Planning assessment	Q3.5		-.847					.885
	Q3.1		-.817					
	Q3.2		-.772					
	Q3.7		-.706					
	Q3.6		-.687					
	Q3.3		-.686					
Using student-involving assessment opportunities	Q3.16			.865				.871
	Q3.17			.839				
	Q3.15			.721				
	Q3.18			.690				
	Q3.14			.436				
Using formal assessment tasks	Q3.22				.826			.848
	Q3.21				.662			
	Q3.23				.639			
	Q3.24				.637			
Using spontaneous assessment opportunities	Q3.13					-.557		.859
	Q3.19					-.554		
	Q3.12					-.501		
	Q3.20					-.477		
	Q3.9					-.353		
Making professional judgements	Q3.30						-.702	.897
	Q3.32						-.673	
	Q3.29						-.500	
	Q3.31						-.477	

During the process of factoring, four items had a lower loading of .30 (Q3.26; Q3.27, Q3.35 and Q3.36); four items were identified as complex variables with the loading of over .30 on more than one factor (Q3.28, Q3.11, Q3.25 and Q3.10); and three items were identified as outlying variables as they were unrelated to other items on the same factor (Q3.33, Q3.34 and Q3.8). Therefore, eleven items were discarded from the

analysis. In the final solution, six factors were successfully extracted, accounting for 71.48% of the total variance ($KMO = .925$, $df = 465$, $p < .001$). Cronbach alpha coefficient for the six factors ranged from .848 to .954, indicating very good internal consistency. Based on the shared commence, the six factors were named: F1 *providing descriptive feedback* (43.20% variance); F2 *planning assessment* (9.05% variance); F3 *using student-involving assessment opportunities* (6.23% variance); F4 *using formal assessment methods* (5.22% variance); F5 *using spontaneous assessment opportunities* (4.39% variance); F6 *making professional judgements* (3.38% variance). Table 4.3 presents the factors, items, factor loading, and alpha coefficients.

The *Practice Scale* was designed to examine teachers' classroom-based assessment practices. Similar to the *Process Scale*, the *Practice Scale* was initially hypothesized to cover seven latent constructs. The results of EFA generated a six-factor solution with 31 items, comprising providing descriptive feedback, planning assessment, using student-involving assessment opportunities, using formal assessment tasks, using spontaneous assessment opportunities, and making professional judgments.

The first factor, with six items, was named the practice of providing descriptive feedback. The items reflected teachers' use of detailed feedback to help students recognize their strengths and weaknesses, find ways of solving problems they have in their learning, and understand what they need to do to improve their work.

The second factor, with seven items, was labeled the practice of planning assessment. It reflected teachers' practices of establishing and sharing instructional objectives and success criteria with students and selecting appropriate assessment methods. Four items were related to the practices of establishing teaching objectives according to the requirements of the curriculum and students' learning needs, and

sharing instructional objectives with students. Two items were about identifying and sharing assessment criteria, and another item was about selecting assessment methods according to students' needs.

The third factor, with five items, was named using student-involving assessment opportunities, which reflected teachers' use of less obvious and instruction-embedded assessment opportunities to collect learning evidence. Two items were concerned with conferences with students or their peers; other items included self-and peer assessment, portfolio assessment, and oral presentation.

The fourth factor, with four items, was interpreted as using formal assessment tasks. Each of the items reflected one type of formal assessment activity used by teachers to collect evidence of students' learning, such as classroom tests, dictation, assignments, and reading aloud and reciting during the lessons. These assessment activities were often made explicit to students.

The fifth factor was named using spontaneous assessment opportunities. The five items were related to assessment opportunities that were incidental or spontaneous and embedded in the instruction. It mainly involved collecting learning evidence through observation and oral questioning.

The sixth factor was concerned with the professional interpretation of assessment information. The four items were all important elements of making professional judgments, such as comparing students' performance against their previous performance or pre-set learning objectives, checking the trustworthiness of judgments, and looking for the overall pattern of students' learning.

4.3 Teachers' conceptions of classroom-based assessment

This section reports EFL teachers' self-reported conceptions of the purposes and process of classroom-based assessment.

4.3.1 Teachers' conceptions of CBA purposes

The *Purpose Scale* asked for teachers' conceptions of CBA purposes. Three types of purposes were identified through EFA: learning improvement, teaching improvement, and student accountability and examination. To explore the extent to which teachers agreed on the three assessment purposes, descriptive statistics were computed. The three purposes are listed in descending order of their mean scores (see Table 4.4).

Table 4.4 Descriptive Statistics of Teachers' Conceptions of CBA Purposes

CBA purposes	Number of items	<i>M</i>	<i>SD</i>	% Agree
Teaching improvement	4	5.121	.630	95.4
Learning improvement	6	4.944	.615	94.9
Student accountability & examination	5	3.663	.888	40.0

Note: % Agree = Percentage of respondents who scored ≥ 4 on the Likert scale

The mean score for the teaching improvement purpose was the highest ($M = 5.121$, $SD = .630$), suggesting that teachers strongly believed that classroom assessment played a significant role in improving teaching. The table also shows the percentage pattern of teachers' conceptions of classroom assessment purposes. A majority of EFL teachers (95.4%) agreed on the teaching improvement purpose. This suggests that most teachers accepted the importance of using assessment information to identify the strengths and weaknesses in their daily teaching and to modify their instruction.

Likewise, the learning improvement purpose had a relatively high mean score ($M = 4.944$, $SD = .615$), suggesting that teachers had strong and positive attitudes towards the learning improvement purpose. In a similar vein, most teachers (94.9%) perceived classroom assessment as important for diagnosing students' learning needs, identifying the strengths and weaknesses in learning, and providing immediate feedback to help improve learning.

In contrast, student accountability and the examination purpose achieved the lowest mean score ($M = 3.663$, $SD = .888$), suggesting that there was little agreement with this purpose. The percentages show that only a small portion of teachers (40.0%) agreed that classroom assessment should focus on learners' final achievement and making students accountable and that teachers' instruction should be planned according to examination requirements to get students well prepared for external examinations. The finding suggests that teachers did not agree with the purpose of using assessment for student accountability and examination preparation.

4.3.2 Teachers' conceptions of CBA processes

The teachers responded to the *Process Scale* where 1 = "Not important at all" and 6 = "Completely important" to indicate their conceptions regarding the importance of CBA processes. Seven factors of CBA processes were identified through EFA: planning assessment, using spontaneous assessment opportunities, using student-involving assessment opportunities, using formal assessment tasks, making professional judgments, providing descriptive feedback, and providing evaluative feedback. The overall features of teachers' conceptions about the classroom assessment process were identified through descriptive statistics.

Table 4.5 Descriptive Statistics of Teachers' Conceptions of CBA Processes

CBA processes	Number of items	<i>M</i>	<i>SD</i>	% High importance
Planning assessment	6	5.066	.549	66.9
Using spontaneous assessment opportunities	6	4.828	.573	49.2
Using student-involving assessment opportunities	6	4.666	.623	41.5
Using formal assessment tasks	5	4.864	.560	52.8
Making professional judgements	3	4.959	.565	74.9
Providing descriptive feedback	7	5.077	.565	75.4
Providing evaluative feedback	4	4.644	.600	42.1

Note: % High importance = Percentage of respondents who scored ≥ 5 on the Likert scale

As Table 4.5 shows, the mean score obtained by planning assessment was relatively high ($M = 5.066$, $SD = .549$), suggesting that EFL teachers believed that planning assessment was important for their students' learning and their own instruction. The percentages show that a majority of teachers (66.9%) placed a high value on planning assessment. The finding suggests that teachers strongly believed that they should establish clear instructional objectives and help students understand those objectives, establish success criteria, and select appropriate assessment methods.

For collecting learning evidence, the mean scores of spontaneous assessment opportunities, student-involving assessment opportunities, and formal assessment tasks were all over 4.0. This suggests that all three types of assessment methods were favored by teachers. Spontaneous assessment opportunities ($M = 4.828$, $SD = .573$) and formal assessment tasks ($M = 4.864$, $SD = .560$) received higher mean scores, whereas student-involving assessment opportunities ($M = 4.666$, $SD = .623$) received a relatively lower mean score. Paired samples t-test statistics with Bonferroni correction were used to

examine whether the differences in mean scores were significant (p value was set at .017, .05/3). The results show that difference between spontaneous assessment opportunities and formal assessment tasks was non-significant ($t = -.951, p = .343$); the differences between student-involving assessment opportunities and spontaneous assessment opportunities ($t = -4.945, p < .017$), and between student-involving assessment opportunities and formal assessment tasks ($t = -4.770, p < .017$), however, were both statistically significant. The finding suggests that the teachers put greater emphasis on both spontaneous assessment opportunities (e.g., teacher questioning and observations) and formal assessment tasks (e.g., classroom tests, textbook exercises, and recitation). In contrast, they showed a relatively lower preference for student-involving assessment opportunities such as self-assessment and peer assessment. The percentages also show that teachers tended to favor spontaneous assessment opportunities (49.2%) and formal assessment tasks (52.8%) to student-involving assessment opportunities (41.5%).

Table 4.6 Comparison of Teachers' Conceptions of Three Assessment Methods

Assessment methods		<i>M</i> difference	<i>SD</i>	<i>t</i>	<i>p</i>
Pair 1	Student-involving assessment opportunities - Spontaneous assessment opportunities	-.162	.459	-4.945	.000
Pair 2	Spontaneous assessment opportunities- Formal assessment tasks	-.035	.520	-.951	.343
Pair 3	Student-involving assessment opportunities- Formal assessment tasks	-.198	.579	-4.770	.000

Making professional judgments received a relatively high mean score ($M = 4.959, SD = .565$), suggesting that teachers believed strongly in the value of professional judgments in improving student learning. The percentages also show that most teachers

(74.9%) acknowledged the importance of making professional judgments. This suggests that the teachers believed that comparing students' performance against the learning objectives and their previous learning performance was imperative for learning improvement.

For providing feedback, the table shows that descriptive feedback had a high mean score ($M = 5.077$, $SD = .565$), whereas evaluative feedback ($M = 4.644$, $SD = .600$) had a relatively low mean score. The result of the paired samples t-test indicates that the differences between the mean scores of the two feedback types were statistically significant ($t = 12.291$, $p = .000$). This suggests that the teachers strongly believed in the value of descriptive feedback for enhancing students' learning, but considered evaluative feedback as less important.

4.3.3 Summary

In this section, the results of teachers' conceptions of CBA are reported. In general, the teachers had strong beliefs that the purpose of assessment was to improve teaching and learning, whereas they slightly agreed with the purpose of assessment being to make students accountable and getting students prepared for examinations. It appeared that the teachers strongly believed that, in relation to CBA processes, planning assessment was valuable for students. They responded that it was important to establish clear instructional objectives, share learning objectives and success criteria with students, and to select appropriate assessment methods. While they emphasized the value of using multiple assessment methods for collecting learning evidence, they showed a higher preference for spontaneous assessment opportunities and formal assessment tasks than for involving students in assessment opportunities. Making professional judgments was also considered by the teachers as imperative for improving students' learning and

agreed that it was necessary to conduct criterion-referenced assessment and pupil-referenced assessment and to ensure that judgments were reliable. The teachers placed a higher value on descriptive feedback than evaluative feedback when asked about the place of feedback.

4.4 Effect of teacher attributes and school environments on teachers' conceptions of CBA

This section reports the results for teachers' responses about the relative effects of teacher attributes (teaching experience, previous education in assessment) and school environments (school type and grade level) on their conceptions of classroom-based assessment. Independent samples t-tests and one-way ANOVA were conducted. Based on Huberman's (1989) study on teachers' professional life cycle, the independent variable *teaching experience* in this study was reorganized into three subgroups: little experience (less than 5 years) (N = 66), medium experience (6 to 20 years) (N = 102) and high experience (over 20 years) (N = 27). The independent variable *previous education in assessment* included two subgroups: teachers who had received education in assessment (N = 126) and those who did not have such kind of experience (N = 69). Regarding *grade level*, since some teachers taught more than one grade level, this variable was coded as two subgroups: lower grade level (N = 84), covering respondents who were teaching only Grade 1, Grade 2, Grade 3 or Grade 4, and higher grade level (N = 81), covering respondents who were teaching only Grade 5 or Grade 6. A further 30 teachers who were teaching both lower and higher grade levels were not included in the analyses. The independent variable *school type* also included two subgroups: public school (N = 172) and private school (N = 23). The effects on teachers' conceptions of CBA purposes and CBA processes are reported separately.

4.4.1 Effect on teachers' conceptions of CBA purposes

4.4.1.1 Effect of teaching experience

Table 4.7 shows conceptions of CBA purposes reported by three groups of teachers (Little, Medium and High experience). It shows that all three CBA purposes had the highest mean score among teachers with little experience. Teaching improvement ($M = 5.037$, $SD = .691$) and learning improvement ($M = 4.921$, $SD = .637$) had the lowest mean score for teachers with medium experience, while student accountability and examination ($M = 3.504$, $SD = .958$) had the lowest mean score for teachers with high experience. These findings suggest differences among three groups of teachers' conceptions of CBA purposes.

The results of one-way ANOVAs show that there was non-significant main effect for teaching improvement purpose ($F = 2.055$, $p = .131$), learning improvement purpose ($F = .199$, $p = .820$), as well as student accountability and examination purpose ($F = .541$, $p = .583$). It suggests that teaching experience was not a significant factor influencing teachers' conceptions of CBA purposes.

Table 4.7 Teachers' Conceptions of CBA Purposes by Teaching Experience

CBA purposes	Little experience		Medium experience		High experience		<i>F</i> (2,192)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Teaching improvement	5.235	.568	5.037	.691	5.157	.491	2.055	.131
Learning improvement	4.982	.588	4.921	.637	4.932	.617	.199	.820
Student accountability & examination	3.712	.794	3.675	.930	3.504	.958	.541	.583

4.4.1.2 Effect of previous education in assessment

Table 4.8 displays self-reported conceptions of CBA purposes held by teachers in two groups: teachers who had received education in assessment, and those who had received no education in assessment. It shows that teaching improvement ($M = 5.141$, $SD = .668$) received a higher mean from teachers who had received education in assessment, whereas learning improvement ($M = 4.957$, $SD = .597$) and student accountability and examination purpose ($M = 3.730$, $SD = .862$) received a higher mean score from teachers who had received no education in assessment. This finding shows differences in teachers' conceptions of CBA purposes.

Independent samples t-tests, however, show no statistically significant difference for all the three perceived purposes, with all p values above .05. It indicates that teachers' previous education in assessment did not have an impact on teachers' conceptions of CBA purposes.

Table 4.8 Teachers' Conceptions of CBA Purposes by Previous Education in Assessment

CBA purposes	Assessment education		No assessment education		t	p
	M	SD	M	SD		
Teaching improvement	5.141	.668	5.083	.557	.609	.543
Learning improvement	4.937	.628	4.957	.597	-.217	.829
Student accountability & examination	3.627	.903	3.730	.862	-.777	.438

4.4.1.3 Effect of school type

Table 4.9 presents public school and private school teachers' conceptions of CBA purposes. Teaching improvement ($M = 5.125$, $SD = .645$) and learning improvement ($M = 4.948$, $SD = .628$) purpose had a higher mean score among public school teachers, whereas student accountability and examination purpose ($M = 3.835$, $SD = .824$) had a higher mean score for private school teachers. This finding shows differences between public school and private school teachers' conceptions of CBA purposes.

Independent samples t-tests show that the differences between two groups of teachers' conceptions about all three CBA purposes, however, were not statistically significant. This result leads to the conclusion that school type was not a significant factor influencing teachers' conceptions of CBA purposes.

Table 4.9 Teachers' Conceptions of CBA Purposes by School Type

CBA purposes	Public school		Private school		<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Teaching improvement	5.125	.645	5.087	.520	.271	.786
Learning improvement	4.948	.628	4.913	.524	.253	.774
Student accountability & examination	3.641	.896	3.835	.824	-.984	.326

4.4.1.4 Effect of grade level

Table 4.10 presents lower grade level and higher grade level teachers' self-reported conceptions of CBA purposes. It shows that teachers who were teaching higher grade levels gave higher mean ratings in terms of all three assessment purposes than those teaching lower grade levels. The independent samples t-tests, however, show that the discrepancies between lower grade level and higher grade level teachers' ratings for

teaching improvement ($t = -.457, p = .648$), learning improvement ($t = -.957, p = .340$), and student accountability and examination purpose ($t = -.045, p = .964$) were not statistically significant. This finding suggests that grade level at which teachers were teaching was not a significant factor influencing teachers' conceptions of CBA purposes.

Table 4.10 Teachers' Conceptions of CBA Purposes by Grade Level

CBA purposes	Lower grade level		Higher grade level		<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Teaching improvement	5.095	.660	5.139	.560	-.457	.648
Learning improvement	4.911	.618	4.998	.549	-.957	.340
Student accountability & examination	3.638	.863	3.644	.958	-.045	.964

4.4.2 Effect on teachers' conceptions of CBA processes

4.4.2.1 Effect of teaching experience

Table 4.11 shows three groups of teachers' self-reported conceptions of CBA processes. All the seven perceived components of CBA received the highest mean score among teachers with little teaching experience, while all the components received the lowest mean score among teachers with medium experience. This finding suggests differences among three teacher groups' conceptions of CBA processes.

Table 4.11 Teachers' Conceptions of CBA Processes by Teaching Experience

CBA processes	Little experience		Medium experience		High experience		<i>F</i> (2,192)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Planning assessment	5.101	.563	5.034	.565	5.099	.461	.349	.706
Using spontaneous assessment opportunities	4.990	.595	4.729	.573	4.809	.423	4.329	.014
Using student-involving assessment opportunities	4.773	.639	4.598	.618	4.661	.584	1.587	.207
Using formal assessment tasks	5.030	.5483	4.763	.553	4.837	.535	4.791	.009
Making professional judgments	5.0253	.4972	4.902	.622	5.012	.485	1.096	.336
Providing descriptive feedback	5.063	.536	5.074	.590	5.121	.554	.106	.900
Providing evaluative feedback	4.701	.534	4.601	.632	4.667	.6355	.581	.560

The results of one-way ANOVA show that the differences in teachers' conceptions of spontaneous assessment opportunities ($F = 4.329, p = .014$) and formal assessment tasks ($F = 4.791, p = .009$) were statistically significant. Post hoc Tukey tests show that teachers with little experience (less than 5 years) gave higher mean scores to using spontaneous assessment opportunities ($p < .01$) and formal assessment tasks ($p < .01$), when compared with teachers with medium experience (6 to 20 years). This finding suggests that teachers with little experience placed a higher value on both spontaneous assessment opportunities and formal assessment tasks than did teachers with medium experience.

4.4.2.2 Effect of previous education in assessment

Table 4.12 presents the conceptions of CBA processes of teachers grouped according to whether they had previous education in assessment, or not. Using spontaneous assessment opportunities ($M = 4.859, SD = .608$) and using student-involving assessment opportunities ($M = 4.708, SD = .644$) received a higher mean score from teachers who had received education in assessment. In comparison, the other five CBA components received a higher mean score from teachers with no assessment training or education experience. These scores indicate differences between two groups of teachers in their conceptions of CBA processes.

Independent samples t-tests show that the differences were not statistically significant regarding all the assessment process, which suggests that the teachers' conceptions of CBA processes were not affected by their assessment training and education experience.

Table 4.12 Teachers' Conceptions of CBA Processes by Previous Education in Assessment

CBA processes	Assessment education		No assessment education		<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Planning assessment	5.046	.580	5.101	.491	-.699	.504
Using spontaneous assessment opportunities	4.859	.608	4.773	.501	.997	.320
Using student-involving assessment opportunities	4.708	.644	4.589	.578	1.270	.206
Using formal assessment tasks	4.849	.575	4.890	.535	-.484	.629
Making professional judgments	4.955	.583	4.967	.534	-.132	.895
Providing descriptive feedback	5.065	.586	5.100	.5262	-.410	.682
Providing evaluative feedback	4.615	.6234	4.696	.554	-.897	.371

4.4.2.3 Effect of school type

Table 4.13 shows public school and private school teachers' conceptions of CBA processes. Using student-involving assessment opportunities ($M = 4.670$, $SD = .632$) had a higher mean score among public school teachers, whereas the other six CBA components had a higher mean score among private school teachers. This shows the differences between public school and private school teachers' conceptions about classroom-based assessment processes.

The results of independent samples t-tests show statistically significant differences only regarding teachers' conceptions of formal assessment tasks ($t = -2.054$, $p = .041$). In other words, private school teachers favored formal assessment tasks as more effective for improving students' learning, while public school teachers considered

formal assessment tasks as less effective. Teachers' conceptions of the other CBA components, however, were not affected by school type.

Table 4.13 Teachers' Conceptions of CBA Processes by School Type

CBA processes	Public school		Private school		<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Planning assessment	5.079	.560	4.971	.460	.880	.380
Using spontaneous assessment opportunities	4.817	.565	4.913	.633	-.756	.451
Using student-involving assessment opportunities	4.670	.632	4.638	.561	.230	.818
Using formal assessment tasks	4.834	.561	5.087	.511	-2.054	.041
Making professional judgments	4.954	.584	5.062	.578	-.370	.712
Providing descriptive feedback	5.061	.578	5.193	.443	-1.046	.297
Providing evaluative feedback	4.640	.604	4.674	.576	-.258	.791

4.4.2.4 Effect of grade level

Table 4.14 presents lower grade level and high grade level teachers' self-reported conceptions of CBA processes. It shows that teachers who were teaching lower grade levels gave a higher mean score for using formal assessment methods ($M = 4.871$, $SD = .541$), whereas teachers from higher grade levels gave a higher mean score for the other six CBA components. This finding suggests that there were differences between lower grade level and high grade level teachers' conceptions of CBA processes. Independent samples t-tests, however, show that the differences between two groups' conceptions about all the CBA processes were not statistically significant. This leads to the conclusion that teachers' conceptions of CBA were not influenced by the grade level at which they were teaching.

Table 4.14 Teachers' Conceptions of CBA Processes by Grade Level

CBA processes	Lower grade level		Higher grade level		<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Planning assessment	4.974	.552	5.136	.518	-1.938	.054
Using spontaneous assessment opportunities	4.764	.605	4.817	.498	-.613	.541
Using student-involving assessment opportunities	4.595	.675	4.640	.515	-.477	.634
Using formal assessment tasks	4.871	.541	4.837	.517	.417	.677
Making professional judgments	4.897	.619	5.021	.512	-1.397	.164
Providing descriptive feedback	5.046	.562	5.122	.562	-.866	.388
Providing evaluative feedback	4.580	.613	4.722	.589	-1.516	.132

4.4.3 Summary

This section reports the results regarding the effects of teacher attributes (teaching experience, previous education in assessment) and school environments (school type and grade level) on their self-reported conceptions of classroom-based assessment. The findings suggest that the teachers' conceptions of CBA purposes were not affected markedly by either teacher attributes or school environments. As regards teachers' conceptions of CBA processes, the findings show that teaching experience appeared to influence their conceptions about spontaneous assessment opportunities and formal assessment tasks and that teachers with little experience held stronger beliefs about the importance of both spontaneous assessment opportunities and formal assessment tasks than did teachers with medium teaching experience. School type also appeared to have an impact on teachers' conceptions of formal assessment tasks. Private school teachers reported that formal assessment tasks were more effective methods for enhancing students' learning, while public school teachers said that they found formal assessment

tasks as less effective. Previous education in assessment and grade level had exerted no significant effect on teachers' conceptions of CBA processes.

4.5 Teachers' classroom-based assessment practices

The *Practice Scale* elicited teachers' CBA practices. Six constructs concerning CBA practices were extracted through EFA: planning assessment, using spontaneous assessment opportunities, using student-involving assessment opportunities, using formal assessment tasks, making professional judgments, and providing descriptive feedback. Descriptive statistics were computed to identify the characteristics of teachers' practices of these assessment activities.

Table 4.15 Descriptive Statistics of Teachers' CBA Practices

CBA practices	Number of items	<i>M</i>	<i>SD</i>	% High use
Planning assessment	6	5.226	.547	70.8
Using spontaneous assessment opportunities	7	5.070	.528	65.6
Using student-involving assessment opportunities	5	4.548	.645	30.3
Using formal assessment tasks	4	4.955	.598	61.5
Making professional judgments	5	5.035	.526	74.9
Providing descriptive feedback	4	5.039	.559	71.3

Note: % High use = Percentage of respondents who scored ≥ 5 on the Likert scale

As displayed in Table 4.15, planning assessment received the highest mean score ($M = 5.226$, $SD = .547$), suggesting that, on average, the teachers frequently used planning assessment activities. The percentages also show a high reporting of planning assessment by the majority of teachers (70.8%). It appears that most teachers regularly established and shared instructional objectives with students, identified success criteria,

and also selected appropriate assessment methods according to students' needs to ensure the effectiveness of assessments.

When reporting how they collected learning evidence, the teachers gave relatively high mean scores for spontaneous assessment opportunities ($M = 5.070$, $SD = .528$) and formal assessment tasks ($M = 4.955$, $SD = .598$), while student-involving assessment opportunities was given a lower mean score ($M = 4.548$, $SD = .648$). Paired samples t -tests with Bonferroni correction show that differences between the three assessment methods were statistically significant ($p = .017$, $.05/3$) (see Table 4.16). It indicates that the teachers self-reported a frequent use of spontaneous assessment opportunities, such as oral questioning and observation, to gain a timely understanding of students' learning during the instruction. They also frequently used formal assessment tasks, such as classroom tests, textbook exercises, assignments, and recitation tasks, but only occasionally used student-involving assessment opportunities such as self-and peer assessment and conferences with students. The percentage pattern further confirmed this. A majority of the teachers reported high use of spontaneous assessment (65.6%) and formal assessment (61.5%), whereas only a small portion of the teachers reported high use of student-involving assessment opportunities.

Table 4.16 Comparison of Teachers' Practices of Three Assessment Methods

Assessment methods		<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
		difference			
Pair 1	Student-involving assessment opportunities- Spontaneous assessment opportunities	-.522	.596	-12.240	.000
Pair 2	Spontaneous assessment opportunities- Formal assessment tasks	.115	.542	2.954	.004
Pair 3	Student-involving assessment opportunities- Formal assessment tasks	-.407	.649	-8.767	.000

The mean scores for making professional judgments ($M = 5.039$, $SD = .560$) and providing descriptive feedback ($M = 5.035$, $SD = .525$) were relatively high, above 5.0, suggesting that these two kinds of assessment practices were frequently employed by the teachers. The percentage pattern also shows that a majority of the teachers reported high use of the two assessment practices, 74.9%, and 71.3%, respectively. It would appear that most teachers frequently made judgments of students' performance against their previous performance or the specific learning objectives. They regularly provided feedback to students to identify strengths and weaknesses in their learning and to help students improve their learning.

In summary, the teachers self-reported that they frequently employed the strategies regarding planning assessment, including establishing clear instructional objectives, sharing learning objectives and success criteria with students, and selecting appropriate assessment methods. They used multiple assessment methods, including spontaneous assessment opportunities, student-involving assessment opportunities, and formal assessment tasks to collect evidence of learning. More specifically, the teachers appeared to rely heavily on spontaneous assessment opportunities such as questioning and observation and formal assessment tasks such as textbook exercises, tests, recitation, and assignments, whereas student-involving assessment opportunities were used less frequently. They also constantly made professional judgments by comparing students' current performance against pre-set criteria and students' previous performance and checking the reliability of judgments. They reported frequent use of descriptive feedback to help identify strengths and weaknesses in students' learning and to suggest the next steps for learning.

4.6 Relationships between teachers' conceptions and practices regarding classroom-based assessment

This section reports the results of the analysis of the relationships between teachers' conceptions of CBA and their self-reported CBA practices. Correlation analyses were used first to determine if there was a direct relationship between teachers' conceptions and practices. Simultaneous multiple regression analyses were then used to examine how teachers' conceptions predicted their assessment practices.

4.6.1 Relationships between teachers' perceived CBA purposes and their practices

First, the results of correlation analyses show that learning improvement purpose was significantly positively correlated with all CBA practices (see Table 4.17), which suggests that, when EFL teachers held strong beliefs in the purpose of using assessment to help improve students' learning, they were likely to employ all CBA practices. The results also show a significantly positive correlation between teaching improvement purposes and all CBA practices with the exception of using student-involving assessment opportunities to collect learning evidence. This indicates that the teachers, who reported strong conceptions that assessment should be used for the improvement of teaching, frequently implemented most of the CBA practices. The student accountability and examination purpose was significantly positively correlated only with using formal assessment tasks suggesting that the teachers, who believed that it was important to use assessment information to make students accountable, were more likely to use more formal assessment tasks. Because of the positive relationships identified, the three perceived CBA purposes were retained in subsequent regression analyses as possible predictors of teachers' CBA practices.

Table 4.17 Correlations of Teachers' Perceived CBA Purposes and their Practices

	Classroom-based assessment practices					
	Planning assessment	Using spontaneous assessment opportunities	Using student-involving assessment opportunities	Using formal assessment tasks	Making professional judgments	Providing descriptive feedback
<i>Perceived CBA purposes</i>						
Learning improvement	.397**	.413**	.193**	.275**	.406**	.386**
Teaching improvement	.299**	.359**	.076	.171*	.302**	.360**
Students accountability	.022	.088	.082	.203**	.083	.024

Note: ** = $p < .01$, * = $p < .05$

Table 4.18 Predictions of Teachers' Perceived CBA Purposes on their Practices

Predictors	Teachers' CBA practices					
	Planning assessment	Using spontaneous assessment opportunities	Using student-involving assessment opportunities	Using formal assessment tasks	Making professional judgments	Providing descriptive feedback
Beta (β)						
<i>Perceived CBA purposes</i>						
Learning improvement	.384***	.314***	.221*	.223*	.372***	.304***
Teaching improvement	.086	.172*	-.060	.017	.086	.193*
Student accountability & examination	-.117	-.011	.025	.128	-.053	-.105
R^2	.176	.190	.040	.090	.173	.184
adjusted R^2	.163	.177	.025	.076	.160	.171
df	(3,191)	(3,191)	(3,191)	(3,191)	(3,191)	(3,191)
F	13.579***	14.919***	2.679*	6.312***	13.087***	14.384***

Note: *** = $p < .001$, * = $p < .05$

To examine the predictive effect of teachers' conceptions of CBA purposes on their practices, six linear regression analyses were conducted. For each regression analysis, each of the six CBA practices was set as the dependent variable. The predictor variables were learning improvement purpose, teaching improvement purpose, student accountability and examination purpose.

As presented in Table 4.18, the first regression analysis shows a significant model for the practice of planning assessment, $R^2 = .176$, adjusted $R^2 = .163$, $df = (3, 191)$, $F = 13.579$, $p < .001$. The beta weights show that learning improvement ($\beta = .384$, $p < .001$) was a positive predictor of the practice of planning assessment. In other words, when teachers held strong beliefs in the purpose of using assessment to improve learning, they were more like to set clear instructional objectives, share learning objectives and success criteria with students, as well as select appropriate assessment methods.

A significant model emerged for the practice of using spontaneous assessment opportunities, $R^2 = .190$, adjusted $R^2 = .177$, $df = (3, 191)$, $F = 14.919$, $p < .001$. The results of beta weights show that using spontaneous assessment opportunities was significantly predicted by the learning improvement purpose ($\beta = .314$, $p < .001$), followed by the teaching improvement purpose ($\beta = .172$, $p < .05$). This finding suggests that when teachers agreed with the purpose of using assessment for teaching and learning improvement purposes, they were likely to use spontaneous assessment opportunities to collect learning evidence.

The third regression analysis shows that the overall model significantly predicted the practice of using student-involving assessment opportunities, $R^2 = .040$, adjusted $R^2 = .025$, $df = (3, 191)$, $F = 2.679$, $p < .05$. The beta weights show that the use of student-involving assessment opportunities was positively predicted by the learning

improvement purpose ($\beta = .221, p < .05$). In other words, teachers with positive attitudes towards the purpose of assessment being for learning improvement were more likely to use student-involving assessment opportunities such as self-and peer assessment.

The fourth regression analysis shows a significant model for the practice of using formal assessment tasks, $R^2 = .090$, adjusted $R^2 = .076$, $df = (3, 191)$, $F = 6.312$, $p < .001$. The beta weights indicate that using formal assessment tasks was significantly predicted by the learning improvement purpose ($\beta = .223, p < .05$). This suggests that teachers with strong beliefs in the purpose of assessment for learning improvement might use formal assessment tasks frequently.

The fifth regression analysis shows a significant model for the practice of making professional judgments, $R^2 = .173$, adjusted $R^2 = .160$, $df = (3, 191)$, $F = 13.087$, $p < .001$. The beta results indicate that making professional judgments was positively predicted by the purpose of learning improvement ($\beta = .372, p < .001$). It was likely that teachers who identified the importance of assessment for learning improvement would base their decision on professional judgments.

A significant model also emerged for the practice of providing descriptive feedback, $R^2 = .184$, adjusted $R^2 = .171$, $df = (3, 191)$, $F = 14.384$, $p < .001$. The beta weights show that providing descriptive feedback was significantly predicted by the learning improvement purpose ($\beta = .304, p < .001$) and teaching improvement purpose ($\beta = .193, p < .05$). When teachers believed that assessment should be used for adjusting teaching and learning, they were very likely to provide descriptive feedback to students.

The results of multiple regression analyses identify a significant association between teachers' conceptions of CBA purposes and their CBA practices; the learning improvement purpose, in particular, was positively correlated with all six CBA practices.

4.6.2 Relationships between teachers' perceived CBA processes and their practices

The results of correlation analyses show that each of the seven perceived CBA components was significantly positively correlated with all the six CBA practices (correlation coefficients ranged from .197 to .589) (see Table 4.19). Given the positive relationships, the seven perceived CBA were retained in regression analysis as possible predictors of teachers' CBA practices. The predictive effect of teachers' conceptions of CBA processes on their CBA practices was tested through six multiple regression analyses (see Table 4.20). The dependent variable in each analysis was each of the respective six CBA practices. Predictor variables in each analysis were the seven perceived CBA components.

Table 4.19 Correlations of Teachers' Perceived CBA processes and their Practices

	Teachers' CBA practices					
	Planning assessment	Using spontaneous assessment opportunities	Using student-involving assessment opportunities	Using formal assessment tasks	Making professional judgments	Providing descriptive feedback
<i>Perceived CBA processes</i>						
Planning assessment	.516**	.446**	.188**	.275**	.424**	.390**
Using spontaneous assessment opportunities	.366**	.447**	.305**	.296**	.362**	.390**
Using student-involving assessment opportunities	.384**	.453**	.420**	.263**	.368**	.439**
Using formal assessment tasks	.337**	.418**	.230**	.475**	.388**	.425**
Making professional judgments	.391**	.415**	.256**	.329**	.498**	.456**
Providing descriptive feedback	.472**	.526**	.197**	.325**	.487**	.589**
Providing evaluative feedback	.291**	.336**	.217**	.312**	.382**	.388**

Note: ** = $p < .01$

Table 4.20 Predictions of Teachers' Perceived CBA Processes on CBA practices

Predictors	Teachers' CBA practices					
	Planning assessment	Using spontaneous assessment opportunities	Using student-involving assessment opportunities	Using formal assessment tasks	Making professional judgments	Providing descriptive feedback
Beta (β)						
Perceived CBA processes						
Planning assessment	.369***	.095	-.141	-.012	.088	-.065
Using spontaneous assessment opportunities	-.048	.066	.036	-.003	-.050	-.077
Using student-involving assessment opportunities	.086	.157	.437***	-.024	.045	.199
Using formal assessment tasks	.028	.092	.014	.417***	.076	.077
Making professional judgments	.005	-.011	.062	.108	.253**	.052
Providing descriptive feedback	.273**	.349***	-.090	-.008	.201*	.491***
Providing evaluative feedback	-.120	-.087	.109	.047	.026	-.004
R^2	.311	.331	.193	.235	.299	.376
adjusted R^2	.286	.306	.163	.206	.273	.352
df	(7,187)	(7,187)	(7,187)	(7,187)	(7,187)	(7,187)
F	12.078***	13.225***	6.398***	9.191***	11.410***	26.071***

Note: *** = $p < .001$, ** = $p < .01$, * = $p < .05$

The results of the first regression analysis show a significant model for the practice of planning assessment, $R^2 = .311$, adjusted $R^2 = .286$, $df = (7,187)$, $F = 12.078$, $p < .001$. The beta weights demonstrate that teachers' conceptions of planning assessment ($\beta = .369$, $p < .001$) and providing descriptive feedback ($\beta = .273$, $p < .01$) were two positive predictors of planning assessment practices. In other words, when the teachers emphasized the importance of planning assessment, they were more likely to plan assessments. Likewise, when the teachers placed a high level of importance on providing detailed feedback to support students' learning, they would frequently conduct planning assessment activities.

The results of the second regression analysis show a significant model for the practice of using spontaneous assessment opportunities, $R^2 = .331$, adjusted $R^2 = .306$, $df = (7,187)$, $F = 13.225$, $p < .001$. The beta weights demonstrate that teachers' conceptions of providing descriptive feedback ($\beta = .349$, $p < .001$) was a positive predictor of their practice of using spontaneous assessment opportunities. It suggests that when the teachers attached importance to providing descriptive feedback for the improvement of student learning, they reported they would frequently use spontaneous assessments to collect learning evidence.

A significant model for the practice of using student-involving assessment opportunities also emerged, $R^2 = .232$, adjusted $R^2 = .193$, $df = (7,187)$, $F = 6.398$, $p < .001$. The beta weights show that teachers' conceptions of student-involving assessment opportunities ($\beta = .437$, $p < .001$) was a significant positive predictor of their practices of using such assessment practices. That is, when the teachers placed a high level of importance on student-involving assessment opportunities, they were likely to use such assessments frequently.

The results of the fourth regression analysis show a significant model for the practice of using formal assessment methods, $R^2 = .235$, adjusted $R^2 = .206$, $df = (7,187)$, $F = 9.191$, $p < .001$. The beta weights demonstrate that teachers' conceptions of formal assessment tasks ($\beta = .417$, $p < .001$) was a positive predictor of their practices of using formal assessment tasks. This suggests that teachers' attitudes towards formal assessment tasks were an important factor influencing their self-reported use of formal assessments.

The results of the fifth regression analysis show that the overall model significantly predicted the practice of making professional judgments, $R^2 = .299$, adjusted $R^2 = .273$, $df = (7,187)$, $F = 11.410$, $p < .001$. The beta weights show that teachers' practices of making professional judgments were not only significantly predicted by their conceptions of professional judgments ($\beta = .253$, $p < .01$) but also their conceptions of providing descriptive feedback ($\beta = .201$, $p < .05$). It suggests that when the teachers placed a higher level of importance on professional judgments, they reported they would more frequently make professional judgments when interpreting students' learning evidence. Similarly, when they placed a higher level of importance on the provision of descriptive feedback for students, they reported making professional judgments.

The results of the final regression analysis show that the overall model significantly predicted the practice of providing descriptive feedback, $R^2 = .376$, adjusted $R^2 = .352$, $df = (7,187)$, $F = 23.044$, $p < .001$. The beta weights demonstrate that teachers' conceptions of descriptive feedback ($\beta = .491$, $p < .001$) was a positive predictor of their practices of providing detailed feedback for promoting students' learning. It suggests that the teachers, who placed a higher level of importance on

descriptive feedback, reported frequently providing such feedback to help improve students' learning.

4.6.3 Summary

This section presents the findings regarding the relationships between teachers' self-reported conceptions of classroom-based assessment and their self-reported assessment practices. The results of correlation analyses and multiple regression analyses demonstrate a strong relationship between teachers' conceptions of learning improvement purpose and all the assessment practices, including planning assessment, using spontaneous assessment opportunities, using student-involving assessment opportunities, using formal assessment tasks, making professional judgments and providing descriptive feedback.

The results also suggest strong relationships between teachers' conceptions of CBA processes and their assessment practices. More specifically, teachers' conceptions of planning assessment was an important predictor of their practices of planning assessment. The teachers' conceptions of using spontaneous assessment opportunities was a significant predictor of their practice of using spontaneous assessments. The teachers' conceptions of student-involving assessment opportunities was also a significant predictor of their practice of using student-involving assessments. Their conceptions of formal assessment tasks was a significant predictor of their practices of using formal assessment tasks; their conceptions of professional judgments also influenced their practices of making professional judgments. Finally, the teachers' conceptions of descriptive feedback was a significant predictor for the practices of using descriptive feedback to improve student learning.

Taken together, the findings of this study have confirmed the significant role of teachers' CBA reported conceptions in the implementation of CBA practices.

4.7 Chapter summary

In this chapter, the findings regarding four research questions are reported. The findings on teachers' CBA conceptions show that the teachers had strong positive attitudes towards the purpose of improving teaching and learning, while they slightly agreed with the purpose of making students accountable and getting students prepared for external examinations. The results also suggest that the teachers perceived planning assessment as an important component of classroom assessment. While the teachers placed a high value on using multiple assessment methods to collect learning evidence, they showed a stronger preference for spontaneous and formal assessment methods than student-involving assessment opportunities. The teachers considered making professional judgments as important for learning improvement and also favored descriptive feedback to evaluative feedback.

The results regarding the effects of teacher attributes and school environments on their CBA conceptions show that teaching experience and school type had a significant impact on the teachers' CBA conceptions. In particular, classroom teaching experience had an influential impact on teachers' conceptions regarding using spontaneous assessment opportunities and formal assessment tasks. Their conceptions of formal assessment tasks were also influenced by the type of school at which they taught. Their conceptions, however, did not appear to be affected by previous education in assessment or grade level at which they taught.

When reporting on CBA practices, the teachers self-reported that they frequently conducted planning assessment activities such as establishing clear instructional

objectives, communicating the objectives and success criteria to students, and selecting appropriate assessment methods. Although multiple assessment methods were used by the teachers, spontaneous assessments and formal assessments were identified as being used more frequently; in contrast, student-involving assessments such as self-and peer assessment were employed less frequently. The teachers also reported making professional judgments and providing descriptive feedback to support students' learning frequently.

The results also suggest that there were significant relationships between teachers' conceptions of the purposes and processes of CBA and their CBA practices. That is, the teachers' conceptions of CBA purposes and processes were found to be influential factors on their self-reported CBA practices.

CHAPTER FIVE

RESULTS OF THE CASE STUDY

5.1 Chapter overview

This chapter reports on the results of the case study. The teachers' CBA practices (based on observational and document data) were examined in terms of four dimensions: planning assessment; collecting evidence about students' learning, making professional judgments, and providing feedback. The rationales for teachers' CBA practices are presented following each dimension, which were explored through semi-structured interviews.

5.2 Planning assessment

The findings indicate that instructional objectives were not guided by the national curriculum standards, nor were the learning objectives and success criteria clearly communicated to the students. While the teachers took account of students' learning needs when selecting assessment methods, when planning for assessment, their practice was also heavily influenced the school's assessment policy.

5.2.1 Instructional objectives not guided by curriculum standards

The analysis of the qualitative data obtained from the three teachers' lesson plans shows that the teachers unpacked the national curriculum standards into their own instructional objectives. All teachers had a teaching plan, which outlined the instructional process for a whole semester, designed collaboratively with English teachers teaching the same grade level in their schools. The *instructional objectives* was an important section, where the learning objectives for a whole term, each unit and lesson were presented. The overall objective for a term was established based on the requirements of the national

curriculum standards, thus being relatively general. The objectives, at unit or lesson level, were relatively more detailed, describing the specific language skills, knowledge, and affective attitudes that students were expected to obtain. An extract taken from Amy's teaching plan provides an example of the instructional objectives at term-level, unit-level, and lesson-level, respectively (see Appendix 8).

The interview data, however, shows that the teachers carefully designed the teaching plans to meet the school's requirement rather than, specifically, guide their own instruction. Both Amy and Doris stated that, when designing the teaching plans, they had to set the instructional objectives according to the curriculum standards as required by the school administrators. They confessed that they seldom taught according to their teaching plans; instead, their instruction was primarily guided by the textbooks. As Amy indicated, textbooks "*specified the requirements for vocabulary*" (Amy, Int). Doris also explained that "*textbooks were the major source for the instructional objectives*" in her teaching (Doris, Int#).

It is concluded, therefore, that the instructional objectives were not guided by the curriculum standards. When asked for the reasons, the teachers explained that the overall proficiency requirements specified by the curriculum standards were relatively high. Given the relatively low language proficiency of their students, it was difficult for students to meet the requirements. Consequently, the teachers were unlikely to align their instruction with the curriculum standards. As Kathy put it:

The curriculum standards are more appropriate for students from developed cities in China such as Beijing, Shanghai and Guangzhou, since these students have higher language proficiency and can meet the requirements. However, students from less developed cities such as Chongqing are with relatively lower

language proficiency. It is, therefore, difficult for my students to meet these requirements. How can I set the objectives according to the curriculum standards? (Kathy, Int)

5.2.2 Learning objectives not clearly communicated to students

The observational data shows that the teachers did not explicitly communicate the learning objectives of a lesson, or a particular language activity, to students. They usually began their instruction directly, giving no explanation as to what the students were going to learn, and how they would learn. In one lesson, *Let's spell*, for example, students from Amy's class were expected to learn how to spell letter *I* with a vowel /i:/ (Amy, Obs#1). Amy began her lecture with an introduction to a riddle about animals and then moved to teach students the sound of letter *I*. During the whole process, Amy did not give any explanation of the learning objective or discuss it with the students.

On some occasions, the teachers required students to remember or summarize the content they had learned at the end of an activity. For example, in Doris's case, she carried out a chant activity, in which students were taught first to read the chant, then asked to practice reading the chant in pairs, and finally were invited to recite the chant. At the end of the activity, she asked all students to remember what they had learned by emphasizing in a loud voice that "*Boys and girls, please remember all these words*" (Doris, Obs#4). Despite Doris' attempt to require the students to remember the words, it remained unclear whether these students knew what they had learned. In another example, Amy asked students to summarize what they had learned after she had completed the lesson on *Let's Spell*. Most students could not clearly state what they had learned as they replied that they had learned the words big and pig. Since the pronunciation of the letter *I* was learned through a riddle about animals and the learning

objective had not been clarified, it was no surprise that these students had difficulty explaining what they had learned. It suggests that asking students to summarize what they had learned was ineffective in helping students have a clear idea of the learning objective. The following episode illustrates how Amy guided students to summarize what they had learned:

Teacher (T hereafter): *Ok, let's review what we learn this class 我们来复习下我们这节课主要学了什么? (Let's review what we have learned this lesson)*

Students (S hereafter): *Big, pig*

T: *Ok, Xue (A student's name, pseudonym)*

Student 1 (S1 hereafter): *学了 letter I 的发音 (I learned the pronunciation of letter I)*

T: *Letter I 的发音, ok, very good. 有同学给我说老师我们今天学了 pig, yes or no? (Some students said that they had learned the word 'pig', yes or no?)*

S: *No*

T: *Ok, 我们一起来回忆下含有 letter I 的单词, 关上你的书, close your book, 我们一起来复习, letter I 有哪些? ok, one student say a word, 每个同学说一个. (Ok, let's review the words containing letter I. Close your book, let's review together, letter I, what kinds of words can you remember? Ok, one student say a word).*

S1: *Pig*

S2: *Big*

S3: *Six*

S4: *Big*

T: *Any more else?*

S5: *Fish*

S6: *Pink*

S7: *Thirteen, 13*

T: *Ok.*

(Amy, Obs#1)

The interview data shows that teachers' practice of not communicating learning objectives to students was influenced by students' learning needs. The teachers argued that primary school students often had difficulty in understanding the learning objectives at the beginning of a class. In contrast, they believed that asking students to make a summary of what they had learned at the end of a class would be more effective in helping them understand the objectives. As Kathy said:

They are too young. If you explain to them what they are going to achieve in the very beginning, they will wonder what this means as they have not learned it at all. (Kathy, Int)

Doris also emphasized that her students “*showed no interest*” when she introduced the learning objectives at the beginning of a lesson (Doris, Int). She reflected as follows:

I used to tell them what they are going to learn at the very beginning...I present the objectives on the PowerPoint slide. However, I found them showing no

interest in knowing the objectives. Therefore, I tend to ask students to make a summary at the end so that they can better understand the objectives. (Doris, Int)

5.2.3 Attempts made to explain assessment process rather than to clarify success criteria

The observational findings suggest that the teachers almost never communicated success criteria to their students. When students were to be involved in an assessment task such as oral reading and doing textbook exercises, the teachers directly described the assessment task and how it should be completed, but the criteria for successful performance was not clearly articulated. In Amy's case, for example, a recitation task was organized to check whether students had mastered some new words about animals (Amy, Obs#1). Amy told her students only that they should close their book and recite the talk together following her instruction; she did not clarify the features of recitation performance that would be used to make a judgment.

The teachers sometimes used exemplars to model how an oral production task should be completed, hoping that students would understand how dialogue was constructed. They usually used phrases such as 'for example', 'do it like this', and 'just like this' to draw students' attention to the process of the task. No specific criteria about what the teacher was looking for was shared with students, however. In the following episode, for example, the students were asked to make sentences using the new words (Monday, Tuesday, and Wednesday) that they had just learned. They were shown how to make a dialogue in pairs through a teacher-and-student modeling dialogue. After the modeling, students were asked to practice making dialogue in pairs. The teacher then asked two students to construct their own dialogue and commented on their performance.

During the whole process, the teacher did not communicate with students about the criteria for their performance.

T: *Okay, let's work in pairs. For example, you please stand up, I say "Today is Monday", you say?*

S1: *Tomorrow is Tuesday.*

T: *Yes, do it like this. You say today...?*

S1: *Today is Tuesday*

T: *I say "Tomorrow is Wednesday". Understand?*

S: *Yes.*

T: *Work with your desk-mate, go.*

(Students were working in pairs)

T: *This group, you, have a try. Stand up. Hush!*

S1: *Today is Monday.*

S2: *Tomorrow is Tuesday.*

T: *Tuesday, yes. You, now it's your turn. Today is Monday, go! Oh, Today is Tuesday.*

S2: *Today is Tuesday.*

S1: *Tomorrow is Wednesday.*

T: *Wednesday. Okay, what do you think of them? One, two?*

S: *One*

T: *One sticker, one sticker.*

(Doris, Obs#1)

The interview data confirms that the teachers did not share success criteria with students and suggests that such practice was influenced by students' learning attitudes. The teachers indicated that they did not set specific success criteria as higher achieving students usually had positive learning attitudes and would be responsible for their own learning and achieve the learning expectations without the assessment criteria being clarified. The teachers said that low achieving students had negative learning attitudes, and could not perform well in any subject (including Mathematics and Chinese). They believed that these students would feel compelled if teachers insisted on clarifying the specific criteria. As Doris put it:

I do not establish any specific criteria ... In my opinion, excellent students know the criteria for good work by themselves even if teachers do not make it explicit. They are good at all the subjects, including Chinese and Mathematics. They have good learning habits and are good at everything ... Low achieving students are not good at either Chinese or Mathematics. I seem to force them to learn if I set specific criteria. (Doris, Int)

5.2.4 Selection of assessment methods influenced by both students' needs and school' assessment policy

The interview data shows that the teachers took account of students' learning needs and language proficiency when selecting assessment methods. For instance, Doris mentioned that she decided not to use games when students had been enrolled in Grade 4 as they became less interested in games. Both Amy and Kathy emphasized that they would choose appropriate assessment methods according to students' language proficiency. They believed that the overall language proficiency of each class was

different, and thus teachers should choose the most suitable assessment method according to students' proficiency. As Kathy put it:

The overall proficiency of each class is different. For instance, some assessment activities carried out in this class are inappropriate for another class. In this case, you should consider students' proficiency to select appropriate assessment methods. (Kathy, Int)

The selection of assessment methods was also driven by the school's assessment policy. In both Doris' and Kathy's cases, for example, some students were randomly selected and assessed by the school administrator to examine their English language proficiency at the end of the term. The assessment was high-stakes because the students' result was an important indicator of teachers' teaching effectiveness, which would consequently affect their salary. As a result, they tended to choose assessment methods that would be used by the school administrator. As Kathy put it:

*It (the school's assessment policy) works as a guiding policy, which has influenced my teaching. **If I know how they will assess my students, I have to teach accordingly.** For example, if I know that the school administrator emphasizes students' oral reading ability and oral reading was an assessment format, I will pay a lot of attention to oral reading tasks. (Kathy, Int)*

5.3 Collecting evidence of students' learning

The findings indicate that the teachers employed a variety of assessment methods to collect evidence of students' learning; they included: spontaneous assessment opportunities that were unplanned and embedded in daily instruction; student-involving assessment opportunities that were embedded in regular instructional activities and emphasized students' role, such as oral production activities, self-and peer assessment;

and formal assessment tasks such as oral reading, textbook tasks, teacher-constructed exercises, and recitation tasks.

5.3.1 Using spontaneous assessment opportunities

The findings suggest that the teachers frequently used spontaneous assessment opportunities that were embedded in their instruction, including questioning and observation because they believed strongly in the benefits of such assessments. They used questioning to help modify their instruction, provide immediate feedback to students, and stimulate students' participation in learning. They also observed students' performance and behaviors, using the evidence elicited to adjust their instruction accordingly.

5.3.1.1 Questioning

Questioning was the assessment method most widely used by three teachers to gather information about students' learning during the observations. The questions the teachers mainly asked were closed to acquire factual information, or simple comprehension, from the whole class or an individual student. For instance, in one observed lesson, Amy, teaching a story about animals, asked a series of closed questions to see whether the students had understood the meanings of the sentences or phrases (Amy, Obs#2). In another example, Doris asked a series of yes/no and simple questions to check whether students had understood the content of the lesson, a weather report (Doris, Obs#4).

On some occasions, the teachers attempted to encourage more ideas from students by asking open questions; this turned out to be ineffective as students often kept silent, providing no responses. The teachers then had to revert to asking simple questions to promote students' participation in classroom activities. For instance, in another observed lesson, Kathy, asked the question, "*What can you see?*" to encourage students to express

their ideas about a picture (Kathy, Obs#1). When no student responded to this question, Kathy turned to closed questions to stimulate students' active participation such as "Can you see a boy?", "Can you see a girl?" and "Can you see some meat?" (Kathy, Obs#1).

The qualitative analysis of the three teachers' practices suggests that the focus of oral questions varied according to the phase of a lesson. During instruction, the teachers frequently asked questions to see whether students had understood, or were following their instruction, so as to further modify the instruction if required. For instance, the following episode demonstrates how Doris used oral questions on which to base her instruction for a language activity about weather. First, she played a radio recording and then asked students a question to check whether they understood the content of the weather report. As the students answered the question correctly, Doris changed the form of the question to ensure that they had truly mastered the language point. As the answer was correct again, Doris moved to the next language point by asking another question. She replayed the radio recording to help the students produce the sentence correctly and then taught the sentence, word by word, when the students could not give an appropriate answer. It is evident that Doris used oral questions frequently to inform the instructional procedures, such as asking more questions, playing the radio, and offering further instruction.

T: *Boys and girls, listen. (The teacher played the radio)*

T: *So the Robert can make the weather report. **The Robert say it will be? It will be?***

S: *Sunny in Beijing.*

T: *Sunny, yes, sunny, follow me. **It will be sunny in Chongqing?***

S: *In Beijing.*

T: *In Beijing, good, what else? What else? Adam*

S1: *It will be sunny in London*

T: *It will be...the second one is sunny?*

S2: *It will be cloudy in London.*

T: ***London, listen, let's listen.*** (*The teacher played the radio again*)

T: *Okay, you got it. It.*

S: *It*

T: *Will be*

S: *Will be*

T: *It will be sunny, sunny*

S: *Sunny*

T: *Sunny*

S: *Sunny*

T: *In Beijing*

S: *In Beijing*

(Doris, Obs#4)

At the end of an instructional activity, the teachers usually asked a couple of questions to check whether students had mastered the target linguistic knowledge. Based on students' responses, they usually provided immediate feedback to help students reinforce their achievement. In the following episode, for example, Amy taught the key language points in a talk, after which several questions were asked to check whether

students had mastered the words and sentences. Amy provided positive feedback immediately to confirm students' responses, or provide detailed feedback, to help students correct the mistake.

T: *Now, open your book, turn to page 42. Let's read together. Cat, cat, one two begin.*

S: *Cap, cap, card, card, boat, boat, map, map*

T: *Ok, mum, one two begin.*

S: *Mum, where is my boat? Is it in your toy box? No, it isn't.*

T: *Ok, 句子不是很熟 (You are not familiar with the sentence), every group can get one point.*

T: *Ok, follow me, cap, cap*

S: *Cap, cap*

T: *Card, card*

S: *Card, card*

T: *Boat, boat*

S: *Boat, boat*

T: *Map, map*

S: *Map, map*

T: *Mum, where is my boat?*

S: *Mum, where is my boat?*

T: *Is it in your toy box?*

S: *Is it in your toy box?*

T: *No, it isn't*

S: *No, it isn't*

T: *No, it isn't 是肯定回答还是否定? (Is this an affirmative response or a negative response?)*

S: 否定

T: *肯定回答怎么回答? (What is the affirmative response?)*

S: *Yes, it is*

T: *Ok, I will please a student to answer me. 我请个同学来回答我 (I will ask a student to answer my question) Is it in your toy box? Toy, what's meaning?*

S1: *玩具 (toy box)*

T: *玩具 (toy box).OK very good. Toy box, what's meaning?*

S2: *玩具盒 (toy box)*

T: *玩具箱, 玩具盒 (toy box)*

T: *Ok, next one, I will add the difficulty. 我会增加一点难度 (I will ask a more difficult question), I will please a student to answer us. 请同学回答 (I will ask a student to answer my question), Is it in your toy box? 它是在你的玩具盒里吗? 把它变成一个肯定句, 怎么变? 我们上次讲了的 (Is it in your toy box? How to change the sentence into an affirmative one? We have learned this before), who want to try? Zhang*

S1: *Yes it is*

T: *把is it in your toy box 改成它在你的玩具盒里. (You should change the sentence 'Is it in your toy box?' into an affirmative sentence) Ok, please.*

S1: *It is in your toy box.*

T: *Ok, very good. Two points for your group.*

T: *It is in your toy box.*

(Amy, Obs#3)

In addition, questioning was frequently employed by the teachers to stimulate students' participation in learning. The observational data shows that students were awarded stickers, or bonus scores, when the questions generated by teachers were answered correctly, to motivate them to participate actively in classroom activities. For example, in one observed lesson, Kathy asked students what they could see in a picture and attempted to engage them in the activity by emphasizing that "*It is easy to get a point*". The following episode illustrates how Kathy used questioning to stimulate students' participation:

T: *Look at this picture, what can you see? This picture (a picture is on the slide)*

S1: *A bed, and a chair, and a desk*

T: *Ok, one point for you. Ok, she can see a bed, a chair and a desk. What else can you see? **It's easy to get a point.***

S2: *Book, Sam, Sam's mum, a doll.*

T: *Ok, one point for you. Anything else? Where is the book?*

S: *On the bed.*

T: *Whose book is on the bed? Do know whose? This book who put it on the bed? Sam's mum or Sam?*

T: *Let's listen.*

(Kathy, Obs#4)

The interview findings suggest that all three teachers thought it was important to ask spontaneous questions to check if students were following or understanding their instructions. As Amy explained, if students could not answer incidental questions such as “*Do you like apples?*”, she would know that students were not following her instructions carefully, or had difficulty in understanding them (Amy, Int).

The teachers also claimed that questioning was an effective way to measure students' current level of knowledge. As Amy put it:

I ask questions to see whether students can answer them correctly or not. This is my first intention and I think it is very important because it reflects students' attainment during a lesson. (Amy, Int)

Moreover, the teachers emphasized that questioning played an important role in engaging students in the instructional process. They said that, as teacher-and-student interaction was a crucial factor that influenced the effectiveness of teaching, it was necessary to use spontaneous questioning. Correctly answered questions were awarded bonus scores, which stimulated students' participation. As stated by Kathy:

Students at lower grade levels are very active but become less active when they enter higher grade levels. They became shy. They might be listening and writing, but they are unwilling to answer my questions. This would affect my teaching as

teaching and learning are interactive. If they remain silent, I become less enthusiastic about teaching. ...To encourage them to answer my questions, I will give bonus scores. If they answer three questions correctly, they will get one score, if five, then they get two scores. (Kathy, Int)

5.3.1.2 Teacher observation

The observational data shows that the teachers frequently used observation to collect information about students' learning. They usually watched students' reactions, or interactions, while giving instruction on language points, or when students were engaged in classroom activities such as oral reading, textbook exercises, and discussion work. During these instructional processes, students' reactions such as reading in a low voice, keeping silence in response to teachers' questions, what they said to teachers, as well as their interactions with peers, acted as signals to the teachers. Based on the information obtained, the teachers evaluated how well their instruction was going and then decided to ask further questions, make explanations, provide immediate feedback, or modify their instruction. For instance, Doris was giving instruction on the word 'Wednesday' when she noticed that some students got confused with the word 'Wednesday' and 'Monday,' and some even pronounced it as 'Wansday' (Doris, Obs#1). Doris then immediately gave a further explanation of the pronunciation to help the students pronounce it correctly. In one of Kathy's observed lessons, while students were engaged in a group discussion task, Kathy noticed that a lot of students asked about the meaning of the word 'everything', and so she provided an explanation of this word to the whole class (Kathy, Obs#1).

The findings from the interviews suggest that the frequent use of observation was related to teachers' beliefs about the nature of observation. The teachers believed that

observation was an integral part of instruction, through which they could immediately modify their instruction and provide appropriate feedback to students. As Kathy put it:

If you do not observe students' reactions, do you want to teach by yourself?

Teachers should observe students' work during instruction. It is necessary that teachers keep an eye on students while they are giving instruction (Kathy, Int)

5.3.2 Using formal assessment tasks

The findings suggest that the teachers heavily relied on a variety of formal assessment methods, including oral reading, textbook tasks, teacher-constructed exercises, and recitation. These formal assessment tasks mainly served a summative purpose as they were used to check how well students had mastered the linguistic knowledge. The frequent use of formal assessment tasks was influenced by the school's assessment policy, exam pressure, and teachers' conceptions of learning.

5.3.2.1 Oral reading tasks

The observational data shows that the teachers used oral reading tasks frequently to check whether students had mastered words and sentences. Typically, after new words or sentences had been taught, an oral reading task was conducted with an individual student or the whole class to check their achievement. Students were also required to read aloud vocabulary and sentence lists included in textbooks.

The information gained from oral reading tasks was sometimes used to adjust teachers' instruction accordingly. For example, Doris used oral reading tasks to modify her instruction. In one lesson, Doris realized, based on five students' oral reading performance, that students could only read the word 'Wednesday' in a falling tone but failed to read the word in a rising tone. Doris then explained to the students how the word could be read in two different tones, after which another ten students were asked

to read the word orally again. She provided further instruction to correct the pronunciation of the word as she noticed that some students had difficulty in pronouncing the word correctly (Doris, Obs#1).

In most cases, however, teachers made evaluative judgments about students' performance in oral reading tasks by providing evaluative feedback (general praise and scores) after students had completed an oral reading task. Doris, for instance, asked two groups of students to read a chant orally to the whole class. She commented on their performance by offering general praise such as 'good job' and 'super', without giving detailed feedback on their performance (Doris, Obs#3). In Kathy's class, students were asked to read the sentences orally to their group leader and were informed that they would get four points if they read all the four sentences correctly (Kathy, Obs#1).

When asked why oral reading tasks were frequently employed, both Doris and Kathy explained that it was because students' oral reading was assessed by school administrators at the end of the term. According to Kathy, students would be motivated to "*read aloud the words carefully*" through their oral reading performance scores (Kathy, Int). In this way, students could achieve a high level of performance when they were assessed by the school administrator.

Amy measured students' oral reading performance frequently in order to meet the expectations of parents. As she explained, most children's parents had received higher education and thus attached great importance to their children's English language learning. Students' oral reading performance was an indication of their English language ability, which could be reported to their parents. As Amy explained:

If students cannot oral read the words and sentences correctly, I will report their performance to their parents who pay much attention to their children's English learning. (Amy, Int)

5.3.2.2 Textbook tasks

The observational data shows that the teachers also used textbook tasks frequently to gather evidence of students' learning, including 'Listen and tick the right picture', 'Written exercises', 'Read and tick the right answer', 'Matching exercises' and 'Completing the sentence.' Typically, students were given a certain amount of time to complete the exercises in class, after which they were asked to present their answers to the whole class; and their performance evaluated by the teachers.

Textbook tasks served the summative purpose of checking the extent to which students had mastered the words and sentences taught so far. The teachers usually provided feedback by giving scores to individual students or a whole group. In Amy's case, for example, the students were given two minutes to complete a 'Read and Tick' task, in which they ticked the right sentence according to pictures. Individual students were then asked to read their answers to the whole class, and their group would get one point if the answers were correct (Amy, Obs#3). In another example, students in Kathy's class were given two minutes to complete a 'Sentence Completion' task according to a picture. Students, similarly, were asked to read aloud their answers to the whole class. Kathy explained clearly how each item was scored and asked group leaders to keep a record of group members' scores obtained in the task. It is evident that textbook tasks were used mainly to check students' attainment and to score their performance (Kathy, Obs#4).

5.3.2.3 Teacher-constructed exercises

The observational data shows that Kathy sometimes used teacher-constructed exercises to check students' achievement. In one class, for example, Kathy presented four closed questions and one multiple-choice item on PowerPoint slides and gave students three minutes to find the answers according to the textbook. Some students were then asked to read aloud their answers, and their performance scored by the teacher (Kathy, Obs#3).

Kathy, in the interview, expressed a preference for teacher-constructed exercises. As she explained, she designed exercises based on students' textbooks or teachers' handbooks, the purpose of which was to improve students' test-taking skills. She believed that it was necessary to cultivate test-taking skills of Grade 6 students before they entered junior secondary schools where teachers' instruction was characterized by an "*examination-oriented*" tradition (Kathy, Int#).

5.3.2.4 Recitation tasks

The observation data also shows that Amy and Doris sometimes used recitation tasks. After they had taught a talk or a story, they would use recitation tasks to check how well students had mastered the linguistic knowledge. In one observed lesson, for example, Doris taught a chant sentence by sentence; students were given several minutes to read the chant by themselves, and then two students were asked to recite the chant to the whole class. Once the students had completed reciting the chant, Doris provided evaluative feedback with general praise, such as 'good', and drew a star for their group on the blackboard (Doris, Obs#3). In another example, Amy used a recitation task at the beginning of a lecture to check how well students had mastered the talk she taught in the previous lesson. Amy made evaluative judgments such as 'very good' and gave summative scores to students who recited the sentence successfully. She also expressed

her disappointment by saying, “*Did you recite the talk at home?*” in a loud voice when students could not perform well (Amy, Obs#1).

The interview data suggest that the teachers considered recitation tasks an effective way to help students remember what they had learned. They believed that through recitation, students would memorize the text and be aware of the learning objectives of a lesson. As Amy put it:

I am fond of recitation tasks, which is very efficient....During the process of recitation, students will try to think about what they have learned. Therefore, they can have a better understanding of the text and know what they have learned this lesson. (Amy, Int)

5.3.3 Using student-involving assessment opportunities

The findings show that student-involving assessment opportunities were used occasionally to elicit evidence of students’ learning, including oral production activities, and student self-and peer assessment. This type of assessment was typically planned by teachers in advance to be part of daily classroom activities in which students often played an active role. It was used mainly to measure students’ outcomes or to locate problems in learning, rather than suggest next steps for learning.

5.3.3.1 Oral production activities

Oral production activities were sometimes used to elicit information about students’ current level of learning; they included dialogue making and oral presentation. For dialogue making, the teachers typically explained how the dialogue could be constructed through exemplars, students then practiced making dialogue in pairs or in groups, or were asked to demonstrate their performance; teachers or peers then provided feedback to these students. The feedback was provided in the form of general evaluative praise

(good job), grades, or stars, which offered no specific information about students' learning and how their learning could be improved. For example, in one observed lesson, Doris carried out a dialogue making activity to check whether students had mastered several sentences about the weather that had just been taught. A group of four students was asked first to make four sentences to show how a dialogue could be made. After several minutes' preparation in groups, two groups presented their work to the whole class; their performance was evaluated by their peers and the teacher (Doris, Obs#4). In another example, Kathy conducted a dialogue-making activity to check students' attainment of the sentences 'Why are you...?' and 'Because I am...'. (Kathy, Obs#3). One student was asked to work with Kathy to show how the dialogue should be made in pairs, following which students were given time to prepare their own dialogues. Four pairs of students were then invited to make their dialogues, and their performance scored by Kathy.

Amy sometimes used oral presentation, known as 'Three Minutes Before Class', in which one student prepared a topic independently before a lesson and then gave a three minutes' presentation at the beginning of the lesson, on which peers were expected to provide detailed feedback. The observational data shows that the feedback focused mainly on the negative aspects of the presentation. In one example, a boy gave a presentation on "*My favorite sports*". All evaluation by five peers focused on weaknesses such as "*His voice was very low*", "*He did not even open his mouth*", "*He spoke too fast*", "*He did not spell some words with stress*", and "*He was not fluent in the beginning*" (Amy, Obs#3). This example demonstrates that, although Amy attempted to encourage students to provide detailed feedback to help improve students' learning, the feedback was mainly negative.

5.3.3.2 Self-assessment

There is no evidence of self-assessment in the observation data obtained from Doris' and Kathy's classrooms. In the interviews, Doris and Kathy confirmed that they rarely used self-assessment. For example, Doris stated that "*I seldom engaged students in self-assessment...I don't think I have done well in this aspect*" (Doris, Int). She confessed that in some cases, she asked students to self-evaluate their overall performance only after she had finished the lecture. This type of self-evaluation was summative in nature as it only required students to answer the question, "*Have I done a good job?*" Kathy also explained that students were asked occasionally to make summative judgments of their performance over the past month, in the form of grade levels, "A, B, C" (Kathy, Int).

The interview findings suggest that the infrequent use of self-assessment in Doris' and Kathy's classrooms was influenced by teachers' heavy workload. As explained by Doris, she had to complete a whole textbook within one semester, which resulted in a relatively busy teaching schedule—she started her classes from 10 a.m. in the morning and taught until afternoon almost every day. As she was stressed by the heavy workload, it was difficult to implement self-assessment with individual students.

The observation data shows that Amy was the only teacher who attempted to use self-assessment to help students locate specific difficulties in their learning. In one observed lesson, a boy was asked to present his textbook written exercise to the class and evaluate his own writing. The following episode demonstrates how the student was involved in the self-assessment task:

T: 第一个, 你自己来评价下你所写的. (*First, please self-evaluate your writing*)

S1: 我写的不是很好 (*I have not done a good job*)

T: 写的不好, 哪里不好? 你要首先找到自已的问题. (*You did not do a good job, which aspects are you weak in? You should first identify your own problems*)

S1: 顿笔没有到位 (*My copying is inaccurate*)

T: 你还有什么地方, 你看一下你所写的这些字母, 该出头的出了没有? (*Anything else? Look at the letters you wrote, the upper half of some letters needed to be out of the line. Did you write it correctly?*)

S1: 没有 (*No*)

T: *Letter p, letter p.*

(Amy, Obs#1)

As shown in the above episode, this student initially made a summative judgment about his work, stating that “*I have not done a good job*”. With the help of Amy, who raised a question to guide the student to identify specific weaknesses in his writing, the student recognized that his copying was inaccurate. When Amy raised another question to help him pay attention to the letters he had written, the student realized that the letter *p* had not been written in the correct way.

Despite Amy’s attempts to use self-assessment to improve learning, it appears that the potential of self-assessment had not been realized. The findings suggest that she engaged students in self-assessment activities without explaining the criteria for assessing their writing. As a result, students remained unclear about success criteria during the assessment process and, so tended to evaluate their performance in a

summative way. As this episode shows, the boy could only make a summative judgment about his overall performance and had difficulty identifying specific problems in his writing. Moreover, it shows that the positive aspects of students' writing were ignored by the teacher as she guided students to focus mainly on the problems in their writing.

The interview findings suggest that Amy's self-assessment practice was related to her conceptions. On the one hand, she believed self-assessment could be used to help students reflect on their own learning; on the other, she emphasized that identifying specific problems was helpful in enhancing students' learning but paid little attention to their strengths in learning. As she said:

If you think you have not done a good job, you need to locate the specific weaknesses in your learning. This can be set as a kind of criteria against which students' progress can be checked next time. (Amy, Int)

5.3.3.3 Peer assessment

The findings suggest that, in Doris' and Kathy's classrooms, peer assessment was used only occasionally and served mainly a summative purpose. In Doris' case, for instance, peer assessment was conducted when students participated in oral production activities. The judgments made by peers were summative as they mainly took the form of general praise such as 'good' and 'well done'. In one observed lesson, for example, the performance of one group reading a chant aloud together was evaluated by peers from the other three groups. The peers provided mainly evaluative feedback focused on general praise, such as 'good' and 'super' (Doris, Obs#4). In Kathy's class, the students had been asked only occasionally to evaluate their peers' overall performance during the past month, which, in the form of grade levels such as "A, B, C" (Kathy, Int), were summative in nature.

In Amy's class, she attempted to engage students in peer assessment to identify strengths and diagnosis weaknesses in their learning. For instance, in one lesson, Amy encouraged students to make judgments about both the strengths and weaknesses in their peers' textbook written exercises that were presented on PowerPoint slides (Amy, Obs#1). In another observed lesson, students were required to locate the specific problems in a peer's oral presentation given at the beginning of a lesson (Amy, Obs#3).

The interview findings suggest that Amy's practice of using peer assessment was influenced by her conceptions, she strongly believed that peer assessment could help students to reflect on their own learning to both identify strengths and weaknesses. As she explained, "*It encourages students to think about whether they have the same difficulties as their peers, or what their own strengths are*" (Amy, Int). In addition, Amy conceded that peer assessment could stimulate young children's motivation as it offered them an opportunity to be "*a little teacher*", evaluating their peers' work (Amy, Int).

While Amy had attempted to use peer assessment to improve students' learning, its potential had not been fully achieved because the observational data suggests, Amy did not make success criteria explicit to students. After students had completed their work, she immediately involved them in making judgments of their peers' work without any explanation about what constituted the 'quality' of good work. For instance, when a boy had finished his oral presentation, she said: "*Okay, let me ask some students to comment on his performance, who want to try?*" (Amy, Obs#3).

Another observation was that, although Amy guided her students to identify both strengths and weaknesses in peer's work, students tended to ignore the strengths in their peers' work and focused on the weaknesses. In one observed lesson, for example, students were engaged in evaluating peers' textbook written exercises. Although Amy

had stressed that they should identify the strengths first, one student immediately pointed out a weakness in a girl's work by saying, "*The copy writing was not accurate*". Even though Amy attempted to draw students' attention to the strengths of peers' work by restating the word 'Strengths' twice, another student still said, "*We cannot see the word milk clearly*" (Amy, Obs#3).

5.4 Making professional judgments

A lack of academic criteria, against which assessment evidence could be compared when making judgments, was evident in the data. While the pupil-referenced assessment was used occasionally, there was a high frequency of norm-referenced assessment and pupil-referenced assessment. The teachers, however, sometimes checked the reliability of the judgments.

5.4.1 Lacking criterion-referenced assessment

As indicated in the previous section, closed oral questions, oral reading, textbook tasks, teacher-constructed exercises, and other similar methods were usually used by the teachers to see how well students had attained the learning objectives. These techniques were often used in place of academic criteria to make professional judgments about students' achievement. As explained by Kathy, "*the correct answers of closed oral questions were the basis for my judgment*" (Kathy, Int). As students would be assessed based on their own performance, this seemed an effective way to minimize the negative impact of competition. The interview data, however, indicates that the teachers did not set specific criteria against which students had to work. Doris underlined this lack of criteria by stating that "*I seldom establish specified criteria against which students should be assessed*" (Doris, Int).

Kathy also indicated that she did not set any specific requirements for an acceptable level of attainment. As she pointed out:

In theory, it is necessary to establish specific criteria. However, in practice, the criteria are rather flexible due to the changeable classroom circumstances.

(Kathy, Int)

The interview findings suggest that the lack of specific academic criteria was related to students' learning attitudes. Both Kathy and Doris stressed that higher achieving students always had positive learning attitudes and thus would achieve high attainment even though the teachers did not share the specific criteria with them. For lower achieving students with less positive attitudes towards English learning, setting specific criteria could mean they were forced to achieve the standards. As a result, the teachers believed that there was no need to establish specific criteria.

5.4.2 Pupil-referenced assessment occasionally used

The observational findings suggest that the teachers occasionally used students' past performances against which to interpret assessment information. In one lesson, for instance, Amy, in evaluating a student's textbook written work, assessed it as better than previous work. As Amy pointed out, "*the progress had been made regarding the handwriting of letter 'k', which had been written in a wrong way by a lot of students before*" (Amy, Obs#1). In another example, Kathy criticized one student during an oral reading task as he did not even "*open his mouth*" (Kathy Obs#2). This student was later praised by the teacher because "*he has opened his book*" (Kathy, Obs#2) in a following oral reading task. In this case, the student's previous performance was used as a criterion against which his current performance was judged.

The interview findings suggest that the use of pupil-referenced assessment was influenced by teachers' positive conceptions. All three teachers said they believed that pupil-referenced assessment, which allowed comparisons between students' current and past achievement, would help students recognize their learning progress. This type of assessment, they said, would enhance students' motivation, especially lower achieving students. As indicated by Kathy:

We should compare the students' current performance with their past performance. We can point out any progress made by the students. This can encourage them to make more progress in the future. (Kathy, Int)

Amy also reported that the information about students' learning progress could be shared with parents to encourage students to work harder. As she put it:

I will report students' progress to their parents, who will also encourage their children. As a result, students' learning motivation would be enhanced. (Amy, Int)

5.4.3 High frequency of norm-referenced assessment

The observational findings show that, for the three teachers, making judgments with reference to other students' performance was a common practice. They deliberately created a competitive classroom environment by making comparisons between groups or individual students. Amy and Doris, for example, frequently encouraged students from different groups to compete. Students were sitting in four groups, each comprising about ten students, in their classrooms. When students in a group had performed well in a task, they were awarded stickers or bonus scores, thus creating competition among students. In one lesson taken by Doris, for instance, students involved in an oral reading task were told that they were going to compete in groups. All students cheered for their

own groups to get more stickers. This practice had a negative impact on students' social relationships as students from the group who had lost the stickers became disappointed and upset (Doris, Obs#1). A competitive environment was also evident in Kathy's classroom. For instance, at the end of one lecture, after reading aloud students' scores obtained during the past month to the whole class, she asked students to compare their scores with each other. One student was laughed at by his peers when the teacher reported that he had a relatively low score (Kathy, Obs#1).

The observation data also shows that the teachers used general praise frequently and gave stickers or bonus scores to students who achieved well in the task. While these students were rewarded for higher achievement, other students were encouraged to follow their good example, thus encouraging competition. For instance, Amy praised one group because all the students had completed their homework, by saying "*Very good, all group members had completed their homework*". She then asked students who had not completed their homework to stand and told them, in a loud voice, to finish the exercise in class. In this case, the group being praised was set as a model for their peers (Amy, Obs#2).

Another typical evaluative activity that involved comparison among students was to present students' work in front of the class for peers to comment on. Although the major purpose was to identify strengths and weaknesses in students' work through peer evaluation, the peers often focused on weaknesses rather than strengths; for those being assessed, this was discouraging. In one of Amy's observed lessons, two students, commenting on a girl's three-minute oral presentation, only pointed out the weak points, including low voice and lack of eye contact (Amy, Obs#2). For the girl being evaluated, this was evidently discouraging as she kept silent, with her head bowed.

When asked why teachers frequently compared students' performance with each other, they said they believed that lower achieving students could be motivated to make efforts to achieve a higher achievement through comparison. For instance, Kathy stated that students would recognize the "*huge difference*" when asked to compare their scores with each other, thus being motivated to work hard (Kathy, Int).

At the same time, the interview data shows that all three teachers, to a certain extent, expressed negative attitudes towards norm-referenced assessment. They all claimed that it could exert a negative impact on students' self-esteem. As Amy explained:

You had better not compare one student with another. If two students are asked to answer my question, and I praise one student, the other student may feel demotivated. This is because children are psychologically vulnerable. (Amy, Int)

To sum up, despite the teachers' critical attitudes towards norm-referenced assessment, the consciously tended to make comparisons between groups or individual students. Such practice exerted a negative impact on students' social relationships and self-esteem.

5.4.4 Checking reliability of judgments

The observational findings indicate that the teachers sometimes checked the reliability of peer assessment by making a final judgment by themselves, or inviting another reliable student to make a judgment. For instance, in Doris' class, one group was oral reading a chant; students from other groups strongly argued that the performance was only 'good'. Doris made the final judgment by saying that the performance was 'super' as she realized that the peers were entirely dishonest (Dors, Obs#2). In another case, peers had been asked to comment on the quality of three students' textbook written work.

Amy invited another reliable student to make a final judgment of which student was the best (Amy, Obs#1).

When asked why they checked the reliability of judgments, the teachers responded that some students were dishonest when evaluating their peers' performance because they wanted their own group to win more scores. As a result, they believed strongly that teachers should make a final judgment to mitigate students' possible dishonesty. As Doris said:

Some students tend to be dishonest, giving fewer stickers or points to other groups in the hope that their own groups can get more stickers. Therefore, I should make my own decision to ensure fairness. (Doris, Int)

The teachers explained that some students had a limited understanding of their peers' learning and lacked peer evaluation skills. In other words, they might be unable to recognize that their peers had made progress compared to their previous performance. Consequently, the teachers felt that they had to make their own judgments to encourage students, especially lower achieving students, who had made progress. As Kathy pointed out:

Teachers should also make their own judgments. You should not entirely trust students' judgments. For example, for a lower achieving student, I think he has made progress. However, other students might not be able to recognize the progress made by this student. (Kathy, Int)

Doris, similarly, argued that teachers should make their own judgments to enhance lower achieving students' motivation. As she said:

Teachers should ensure that the judgment is reliable. I should encourage those lower achieving students who have performed well this time and give them more bonus scores. (Doris, Int)

5.5 Providing appropriate feedback

Teachers occasionally provided descriptive feedback to identify successful components of attainment and to help specify how errors can be corrected. They also frequently provided evaluative feedback to motivate students' achieve higher achievement using strategies such as rewarding, approving, punishing, and disapproving.

5.5.1 Descriptive feedback occasionally used to identify attainment and specify improvement

Data from the observations show that the teachers occasionally provided descriptive feedback to identify evidence of learning and to identify strategies for improvement. The former was used to identify what had been achieved successfully, while the latter was concerned with where the errors were and how errors could be corrected.

5.5.1.1 Providing descriptive feedback to identify attainment

There was evidence that the teachers occasionally provided descriptive feedback about performance or behavior considered necessary for success. For instance, in one observed lesson, students were engaged in an oral reading task focused on reviewing the words and sentences. Kathy expressed her satisfaction by stating that "*Every child has open his mouth and was reading just now*". It seemed that '*opening your mouth and reading*' was regarded as the successful component of the oral reading task. The students received specific praise when they had achieved this criterion (Kathy, Obs#2).

The teachers also provided feedback to affirm students' achievement through a restatement of the response, thus making teachers' expectations for success explicit. This type of feedback was often followed by extensive exercises to reinforce the achievement. For example, in one of Doris' observed lessons, a student had made a sentence successfully using the learned sentence structure; her performance was affirmed by the teacher by repeating the sentence. Extensive oral reading exercises were also provided to the whole class to reinforce the learning attainment:

T: *This one, Shenzhen,*

S1: *It will be windy in Shenzhen.*

T: ***It will be windy in Shenzhen, very good. Let's read together.***

S: *It will be windy in Shenzhen.*

(Doris, Obs#3)

In another lesson taken by Kathy, a student had difficulty in producing a phrase using the words just learned. With the help of peers, the student provided a response that met the teacher's expectations. Kathy then restated the phrase and engaged the whole class in reading the phrase together to reinforce the knowledge.

T: *Ok, what can you do? 你能为她做什么了? (What can you do for her?)*

S1: *捡起来 (pick it up)*

T: *Ok, 捡起来怎么说? (how to say 'pick it up'?) P-i-c-k, can you read?*

Follow me, pick up

S: *Pick up*

T: *Pick up*

S: *Pick up*

T: *Pick up everything,*

S: *Pick up everything*

(Kathy, Obs#1)

5.5.1.2 Providing descriptive feedback to specific improvement

The observation findings also reveal that the teachers sometimes provided descriptive feedback to identify errors and help students understand how the errors could be corrected. This type of feedback was provided in three forms. First, the teachers pointed out directly what was wrong and engaged students in correction activities, sometimes using oral questions to help students correct the mistakes. For instance, in Amy's case, several students were asked to read aloud a new word they had just learned. Through observation, Amy recognized that the students had difficulty in pronouncing this word correctly. She immediately pointed out the error and asked a question to help students correct the pronunciation (Amy, Obs#4). The teachers also invited peers to help correct the errors. For instance, in Doris' class, some peers pointed out a mistake a student had made while reading aloud a sentence related to a picture. The following episode shows how Doris engaged peers in error correction.

T: *This one, Haikou, you have a try.*

S1: *It will hot in Haikou*

S: *错 (It is wrong), be*

T: *You help him.*

S2: *It will be hot in Haikou.*

T: *Yes. It will be hot in Haikou.*

T: *Group one, very well.*

(Doris, Obs#3)

Second, the students were sometimes engaged in self-checking activities. In one of Amy's lessons, for instance, while students were doing a textbook written exercise, they were asked to self-check their handwriting carefully to see whether they had written letter *K* in a correct way. Since the students had confused the capital and lowercase of this letter, self-checking was used to help students understand that they should improve their handwriting of the letter (Amy, Obs#1).

As well as teacher-directed correction and self-checking, teachers sometimes expressed specific expectations of what needed improvement in students' work. For instance, in Doris' case, one student made a sentence, "*It will be windy in UK*". Doris then pointed out that "*The United Kingdom is a country*" and further stated that "*You can say it will be windy in England*" (Doris, Obs#3). In this example, the student was expected to talk about the weather in a specific area rather than a big country. In another example, after a student in Amy's class successfully had described an animal using the word taught by the teacher, Amy showed approval of the response by saying, "*Ok, this description is good*". Meanwhile, she expressed her expectation for a better response by saying, "*You can say it has big eyes*" (Amy, Obs#2).

The findings above provide evidence of how teachers used descriptive feedback to identify successful components of achievement and to specify how errors can be corrected; this type of feedback, however, was not frequently used by the teachers.

The interview findings suggest that class size, students' learning needs, and teachers' workload affected teachers' practices. For example, the teachers were required

to teach more than one class, and, as the class sizes were large, it was difficult for them to provide detailed feedback to individual students. Both Doris and Kathy indicated that class size had constrained the use of feedback to identify students' strengths and weaknesses. As Kathy said:

I need to teach so many classes. There are so many students in each class. It is impossible for me to provide feedback to modify the learning of every student.

(Kathy, Int)

Students' learning attitudes also impacted on teachers being able to give descriptive feedback. In each class, there were lower achieving students who appeared to hold negative attitudes towards learning. The teachers believed that providing descriptive feedback made no difference to the motivation of these students, and so they tended not to provide descriptive feedback to them. Doris reflected as follows:

There are always some lower achieving students. They have very negative learning attitudes. My only expectation for them is that they can be disciplined during my teaching. (Doris, Int)

In the interview with Kathy, she said:

There are some students with low proficiency and are not disciplined in class. No matter how hard I have tried to provide them with detailed feedback to encourage them, it never works. (Kathy, Int)

The curriculum workload was another factor as teachers indicated that they had to complete an overcrowded syllabus prescribed by their schools. As they had to devote extra time and effort to catch up with the syllabus, they found it difficult to implement appropriate feedback regularly. Both Amy and Doris complained about the heavy workload. As Amy put it:

There are six learning units and two review units, but there are only three lessons each week. Our school has set strict syllabus requirements, and we have to spare no effort to meet the requirements. (Amy, Int)

Doris also explained the stress brought by the heavy workload:

The heavy teaching load is an important problem...I have no time to prepare my lessons carefully. I am very busy this term. My lessons are not well prepared.

(Doris, Int)

5.5.2 Evaluative feedback frequently used to enhance students' motivation

The findings suggest that the teachers frequently provided evaluative feedback, which was categorized into four types: rewarding, punishing, approving, and disapproving.

5.5.2.1 Rewarding

The three teachers frequently used rewarding as a feedback strategy to motivate students who made a great effort in their work or behavior. This type of feedback took a variety of forms and included symbols, treats, recognition of students' performance by the whole class, and bonus scores. For instance, the teachers often used symbols such as stickers and stars to bring fun to feedback. In one lesson by Amy, a boy who had difficulty spelling the word 'six' correctly made an effort to correct the pronunciation with the help of a peer. The teacher then gave a sticker to this student as a reward for his efforts. As she said, "*We still give one sticker to this student as an encouragement*" (Amy, Obs#1).

Treats were used as another kind of reward. As Doris explained, the stars or stickers that students had won during assessment activities were calculated by their group leaders; the students received a homework-free reward when they had won five stars.

Another form of reward was to make students' performance recognized by the whole class by providing students an opportunity to be a teacher. For example, in Doris' case, a girl who had oral read a new word very fluently was given an opportunity to be a little teacher, teaching the whole class to read the word together (Doris, Obs#1).

The teachers also used bonus scores to motivate students. For example, in one observed lesson by Amy, students were asked to spell the letters they had learned in a previous lesson together. Amy awarded each group one bonus point as "*every group has done a good job*" (Amy, Obs#1). In another observed lesson by Kathy, students were asked to review the words and sentences through oral reading tasks at the beginning of a lecture. While the students were engaged in the task, Kathy kept an eye on their performance. Each student was awarded one bonus point due to their good performance. As Kathy stated:

Every student has done a good job. You were looking at the words. All of you can get one bonus point. (Kathy, Obs#3)

5.5.2.2 Punishing

In contrast, punishment was used by the teachers to express their dissatisfaction with students' behavior or attitudes. This type of feedback was provided when the teachers found that students were not concentrating on their learning or had not made sufficient effort in their work. It was often provided through dismissing students from the classroom community or removing rewards. In one observed lesson, for instance, some students, who had not completed their homework, were asked to stand up and were not allowed to sit down till they had completed the homework in front of the whole class (Amy, Obs#2). In this case, *asking students to stand up* was used as a punishment for students' inappropriate learning attitudes. An example of the removal of rewards in

Doris' classroom, one group was not listening to her instructions carefully while she was talking about how to make a sentence using the word they had just learned. Doris gave a warning to this group by saying "Bye-bye" and taking off a sticker (Doris, Obs#1).

5.5.2.3 Approving

Approving, an overall expression of satisfaction, was commonly used by the teachers when they judged that students' attitudes and achievement in work were satisfactory. This type of feedback often led to rewards and typically took the form of using praise such as 'very good', 'well done', 'good job', 'wonderful', and 'she is really good'. For instance, in the following example, Doris asked students to read word cards. She expressed her approval of their performance through praise when noticing that the students had read the word cards fluently and loudly:

T: *Okay, let's read some words, this one?*

S: *Clever*

T: **Good!** *Clever. This one?*

S: *Cool*

T: *Okay, **good!** How about this one?*

S: *Naughty.*

T: *Hush. Okay, you please.*

S1: *Naughty.*

T: *How about this one?*

S1: *Nice.*

T: **Well done, nice.** *Thank you, sit down, please. You.*

S2: *Cute*

T: *Cute, wonderful. You, please*

S3: *Cute.*

T: *This one?*

S3: *Shy.*

T: *Yes, good. You, please*

S4: *Old*

T: *Old, wonderful. Sit down, please. How about this one?*

S5: *Cool.*

T: *Yes, look at me. This one?*

S: *Cute*

S: *Cool*

S: *Clever,*

T: ***Wonderful***, *please remember all the words.*

(Doris, Obs#1)

5.5.2.4 Disapproving

Disapproving, an overall negative type of feedback was provided when the lack of effort or concentration was judged to be the cause of students' poor performance. This type of feedback by the teachers mainly took the form of explicit expressions of anger or disappointment, which aimed to be corrective of students' attitudes or behavior. In one observed lesson, for instance, Amy expressed her disappointment with students who had

not completed a recitation task at home by saying, “*Have you recited the talk at home?*” in a loud voice (Amy, Obs#1). In another case, Doris asked students to practice oral reading four sentences within groups, following which one group was invited to read the sentences together orally. Having noticed that the students could not read the sentences fluently, Doris criticized them in a loud voice by saying, “*You have not practiced oral reading the sentences carefully at all*” (Doris, Obs#3).

In summary, the above findings provide strong evidence that the teachers frequently provided evaluative feedback, including rewarding, punishing, approving, and disapproving. The interview data suggests the frequent use of evaluative feedback was associated with teachers’ beliefs in the value of evaluative feedback in enhancing students’ learning motivation. As Amy indicated:

The stickers or stars that students get will be transferred into prizes and will be awarded to them at the end of a term. Thus, every student will be motivated to work hard to receive a prize. (Amy, Int)

Likewise, Kathy strongly said she believed that rewards and scores were helpful for motivating students to achieve expected performance. As she put it:

Students will be motivated by scores, especially those medium-ranking students because every student has self-esteem. Nobody wants to be lower achieving students. Thus, they will attempt to work hard and achieve good performance. (Kathy, Int.)

In summary, the findings show that the teachers did not frequently provide descriptive feedback to help specify attainment and improvement, which was constrained by class size, students’ language proficiency, and teachers’ heavy workload.

In contrast, the teachers frequently provided evaluative feedback, which was influenced by teachers' strong beliefs that this type of feedback could enhance students' motivation.

5.8 Chapter summary

This chapter has reported the data indicating how teachers implement classroom-based assessment as well as the rationales for their assessment practices. The findings suggest that the teachers did not utilize fully classroom-based assessment to support young EFL learners' learning. First, as far as planning assessment was concerned, the teachers' instruction was not guided by the curriculum standards, and they did not clearly communicate the learning objectives and success criteria to students. Second, while multiple assessment approaches were used to collect evidence of learning (spontaneous assessment opportunities, formal assessment tasks, and student-involving assessment opportunities), not all assessment methods were used to suggest actions that could be taken to improve learning. Spontaneous assessment opportunities such as questioning and observation were constantly used to modify instruction and provide timely feedback to students. The teachers also relied heavily on formal assessment tasks and used these tasks summatively. Student-involving assessment opportunities were rarely used and were mainly used to measure students' learning outcomes or locate problems in learning; they provided little information on the next steps for learning. Third, teachers frequently used norm-referenced assessment, whereas criterion-referenced and pupil-referenced assessments were not used consistently as the basis of making judgments. Fourth, teachers frequently used evaluative feedback to motivate students to achieve higher achievement, such as rewarding, approving, punishing, and disapproving. In contrast, descriptive feedback was used only occasionally to identify successful components of attainment or to help specify how errors could be corrected. A synthesis of the findings, most importantly, suggests that teachers' classroom-based assessment practices were

“mainly teacher directed and controlled” (Carless, 2011, p. 105) and that students’ role was underplayed in the whole process. It was evident that the students did not have a clear idea of where they were going; they were not actively involved in evaluating and monitoring their own learning, and they were not engaged in taking actions to close the learning gap.

The findings suggest that teachers’ attempts to implement classroom-based assessment, to a certain extent, were strengthened by their personal conceptions. Their beliefs in the value of questioning and observation in enhancing learning had led them to use these assessment strategies constantly. Likewise, teachers’ positive attitudes towards self-and peer assessment had resulted in their attempts to engage students in reflecting on their own work.

It has also revealed that teachers’ CBA practices were greatly constrained by teachers’ beliefs about learning, students’ learning needs and attitudes, parents’ expectations, as well as contextual factors such as class size, heavy workload, school’s assessment policy, and examination-driven system. These factors were potential impediments to the successful implementation of CBA.

CHAPTER SIX

GENERAL DISCUSSION

6.1 Chapter overview

The overall aim of this research was to investigate primary school EFL teachers' conceptions and practices regarding classroom-based assessment in China. To facilitate the investigation, an explanatory sequential mixed-methods study was conducted. Chapter Four presented the results of the survey study in relation to teachers' conceptions about classroom assessment, the effect of teacher attributes and work environments on their conceptions, teachers' self-reported classroom-based assessment practices, and the relationships between teachers' conceptions and assessment practices. Chapter Five reported on the results of the case study, which examined EFL teachers' classroom-based assessment practices in authentic classroom settings, and elicited the rationales for their assessment practices. The two chapters, collectively, presented the findings in respect of the following four research questions:

RQ1: What are EFL teachers' conceptions about classroom-based assessment?

RQ1.1 What are EFL teachers' conceptions of the purposes of classroom-based assessment?

RQ1.2 What are EFL teachers' conceptions of the processes of classroom-based assessment?

RQ2: To what extent do teacher attributes (teaching experience and previous education in assessment) and work environments (school type and grade level) influence their conceptions of classroom-based assessment?

RQ3: How do EFL teachers implement classroom-based assessment practices?

RQ4: What are the relationships between teachers' conceptions and practices regarding classroom-based assessment?

The purpose of this chapter is to discuss the major findings of the questions in relation to the existing literature. To guide the discussion, a summary of the findings regarding each research question is first presented, followed by an interpretation of the findings with reference to the literature and the context of the study.

6.2 Teachers' conceptions of classroom-based assessment

The first question concerned teachers' conceptions of classroom-based assessment, the findings of which are discussed from the two perspectives of the purposes and processes of CBA separately.

6.2.1 Teachers' conceptions of CBA purposes

The EFL teachers clearly acknowledged the formative purpose of classroom-based assessment. They strongly emphasized the purpose of using assessment to modify teachers' instruction, identify students' learning needs, and improve students' learning. In contrast, they attached low levels of importance to the purpose of making students accountable through assessment and using assessment to prepare students for examinations. This finding is consistent with previous studies in L2 contexts (Büyükkarcı, 2014; Guadu & Boersma, 2018; Muñoz, Palacio, & Escobar, 2012; Önalın & Karagül, 2018; Rogers et al., 2007; Troudi et al., 2009) as well as general education (Eren, 2013; Postareff, Virtanen, Katajavuori, & Lindblom-Ylänne, 2012; Sach, 2012; Young & Jackman, 2014). All these studies suggest that teachers believe that classroom-based assessment plays a vital role in improving teaching and learning.

One possible explanation for teachers' strong beliefs in the purpose of using assessment to improve teaching and learning is the assessment reform in China. As discussed previously, the Chinese government has tried to improve the assessment system to cope with the unintended washback of high-stakes testing in China. The NECS, a guideline document for English language education in China, stipulates that classroom-based assessment is an integral part of regular instruction (Ministry of Education, 2011). Assessment reform can exert an influential impact on teachers' conceptions of assessment. Davison (2007), for instance, reported that teachers in Hong Kong began to appreciate the benefit of assessment for learning following an assessment reform. It is, therefore, understandable that primary school EFL teachers in the present study placed a high value on the purpose of assessment for improving learning and teaching.

This finding might also reflect the international interest in the assessment of young language learners. Since the start of the 21st century, classroom-based assessment has been advocated as desirable for young learners as it caters to students' individual needs as well as promoting self-regulated learning (Bailey, 2017; Cameron, 2003; Hasselgreen, 2012; McKay, 2006). With the worldwide trend towards CBA for young learners, it is perhaps not surprising that primary school EFL teachers in China show a preference for formative classroom assessment.

Moreover, teachers' strong beliefs in the purpose of assessment for improving teaching and learning might be explained by the relatively negative influence of summative testing. There is substantial research evidence that summative testing can exert a negative impact on classroom instruction in both general education and L2 contexts (Chen & May, 2016; Chik & Besser, 2011; Hoffman, Assaf, & Paris, 2001; Pedulla et al., 2003; Shepard et al., 2018). This line of research has shown that

summative testing often results in the ‘teaching to tests’ or ‘learning to tests’ modes, which have detrimental effects on students’ motivation.

6.2.2 Teachers’ conceptions of CBA processes

The findings from the survey study indicate that the teachers emphasized the importance of planning assessment, collecting evidence through multiple methods, making professional judgments, and providing descriptive feedback to enhance students’ learning.

6.2.2.1 Planning assessment

The EFL teachers, in this study, agreed on the importance of establishing clear instructional objectives, sharing the objectives and success criteria with students, and selecting appropriate assessment methods according to students’ needs. This finding is similar to Büyükkaracı’s (2014) study, in the Turkish EFL context, which reported that primary school EFL teachers had positive beliefs about the value of sharing learning objectives with young learners. While Büyükkaracı’s study focused only on how teachers appreciated the value of sharing the learning objectives, the current study investigated teachers’ conceptions of various aspects of assessment planning, including establishing the learning objectives, sharing success criteria and selecting appropriate methods. In this respect, this study adds to the literature that suggests L2 teachers generally attach great importance to various components of assessment planning.

6.2.2.2 Collecting evidence of students’ learning

The EFL teachers reported in the survey study that they recognized the value of using multiple assessment methods to collect learning evidence. They were positively disposed towards the use of spontaneous assessment opportunities during daily instruction, such as teacher questioning and observation. In addition, they were highly

positive towards using student-involving assessment opportunities, such as self-and peer assessment, portfolios, and oral presentations, on an ongoing basis. They also placed considerable value on formal assessment tasks such as classroom tests, textbook exercises, written assignments, oral reading, and recitation. This finding is supported by other research, which has reported that L2 teachers hold positive attitudes towards the use of multiple assessment methods to promote learning (Chan, 2008; Inbar-Lourie & Donitsa-Schmidt, 2009; Shohamy et al., 2008; Troudi et al., 2009).

The teachers' self-reported strong beliefs in the value of using multiple assessment methods might be associated with the need to cater to students' varying language proficiency and learning needs in China. In primary schools in China, teachers usually have to teach large classes with up to 50 students (Wang, 2009). Moreover, a majority of teachers often teach more than one class as well as more than one grade level. In these circumstances, teachers have to cope with students with varying English language proficiency. It is possible that the teachers in the current study expressed favorable attitudes towards a variety of assessment methods because multiple assessment methods have been advocated as enabling teachers to respond to various learning needs (Leung, 2005; Rea-Dickins, 2001; Turner & Purpura, 2016).

Another possible explanation for teachers' positive attitudes towards the use of multiple assessment methods is the characteristics of young language learners. Researchers have argued that it is important for teachers to take account of young learners' unique cognitive, social, and emotional characteristics (Jalongo, 2000; McKay, 2006). For example, teachers should be cognizant of students' interactions and give appropriate feedback in a timely way. Self-and peer assessment is necessary to help young learners take responsibility for their own learning, and small group tasks are

preferred to develop students' social skills such as expressing praise and learning to cooperate and share with others.

Despite teachers' generally positive attitudes towards multiple assessment methods, their conceptions of the three different types of assessment varied. While spontaneous assessment opportunities and formal assessment tasks were strongly favored by the majority of the teachers, student-involving assessment opportunities were perceived as less important. The EFL teachers' favorable attitudes towards spontaneous assessment opportunities (e.g., informal questioning and observation) are encouraging as they can have a role in monitoring their daily instruction and improving students' learning (Turner & Purpura, 2016; Yang, 2012).

Second, the finding that the EFL teachers acknowledged the value of formal assessment tasks has been reported in the literature (e.g., Al-Nouh, Taqi, & Abdul-Kareem, 2014; Brumen & Cagran, 2011). As primary school EFL teachers in China are teaching large-size classes, it is understandable that the teachers in this study placed a high value on formal assessment tasks, which are particularly efficient and objective for large-size classes (Cheng & Sun, 2015).

Third, the finding that teachers had reservations about the value of assessment opportunities involving students (e.g., self-assessment, peer assessment, portfolio) has also been reported in previous studies (Al-Nouh et al., 2014; Sach, 2012). The teachers' negative attitudes may be related to their lack of trust in the validity and reliability of student-involving assessment opportunities. As Inbar-Lourie and Donitsa-Schmidt's study (2009) suggested, teachers tended to question the psychological properties of student-involving assessments, having doubts about whether these assessments could measure students' knowledge accurately.

6.2.2.3 Making professional judgments

The teachers believed that it was important to compare students' performance against the learning objectives and their previous learning progress. These conceptions appear to be shared by researchers who advocate criterion-referenced assessment and pupil-referenced assessment, such as Airasian and Abrams (2003) and Jacobs and Renandya (2019). Airasian and Abrams (2003) claim that by using criterion-referenced assessment the undesirable negative impact of competition can be avoided; students' confidence and motivation can be enhanced if they are assessed on the basis of the learning objectives and not the performance of their peers. Jacobs and Renandya (2019) argue that pupil-referenced assessment is beneficial to self-regulated learning since it helps diagnose students' own learning needs and identify resources and strategies for learning improvement.

The teachers' positive attitudes towards making professional judgments may be explained by the characteristics of young language learners. As young language learners are particularly vulnerable to criticism or feelings of failure, their self-esteem can be strongly influenced by assessment experiences (Bailey & Osipova, 2016; Butler, 2019). Success and a sense of progression to enhance their enthusiasm for learning are, therefore, essential for young learners. McKay (2006) asserts that it is crucial to compare children's progress with specific standards or their past performance to help them gain a sense of achievement. It is not surprising, therefore, that the primary school EFL teachers in this study held positive attitudes towards criterion-referenced assessment and pupil-referenced assessment.

6.2.2.4 Providing appropriate feedback

The findings, consistent with the literature, suggest that the teachers had strong beliefs about the value of descriptive feedback to help students understand what is necessary for achievement and how to overcome difficulties in learning; and for students to be engaged in the closing of their learning gaps (e.g., Brumen & Cagran, 2011; Büyükkarcı, 2014). In contrast, the teachers did not place a high value on evaluative feedback. In Brumen and Cagran's (2011) study, the EFL teachers from three European countries (Czech Republic, Slovenia, and Croatia) were reported to believe that young foreign language learners should be provided with more descriptive and individual feedback rather than with numerical grades. Similarly, Büyükkarcı (2014) found that primary school EFL teachers in Turkey agreed that teachers should provide oral and written feedback about students' strengths and weaknesses, rather than grades and marks.

Such finding is encouraging as research has demonstrated the benefits of descriptive feedback in helping students be aware of the gap between their actual performance and the expected learning objectives as well as understand how to close the gap (Black & Wiliam, 1998a; Brookhart, 2001; Sadler, 1989). Because of their unique cognitive, social, and emotional characteristics, young learners benefit from detailed and descriptive feedback on their learning progress (Brumen & Cagran, 2011).

In summary, the findings of this study suggest that Chinese primary school EFL teachers place a high value on classroom-based assessment. This is in line with Büyükkarcı's (2014) and Young and Jackman's (2014) findings of teachers' strong and positive attitudes towards a range of formative assessment processes, including sharing learning goals, using self- and peer assessment, and providing quality feedback.

6.3 Influence of teacher attributes and work environments on teachers' conceptions of classroom-based assessment

The findings of the survey study suggest that years of teaching experience, as well as the type of school at which teachers were teaching, had a significant impact on primary school EFL teachers' conceptions of classroom-based assessment of young learners. Neither the teachers' previous education in assessment, nor the grade level at which they taught, was identified as an influential impact on their conceptions.

6.3.1 Effect of teaching experience

The findings suggest that teaching experience had an impact on teachers' beliefs about opportunities to collect learning evidence through spontaneous assessments and formal assessments. First, teachers with little experience (less than 5 years) put a relatively high value on spontaneous assessment opportunities, such as questioning and observation, while teachers with medium experience (6 to 20 years) perceived spontaneous assessments as ineffective. The finding that less experienced, or novice, teachers held strong beliefs about the value of spontaneous assessments is consistent with data from Al-Noul et al.'s (2014) study, which indicated that novice primary school EFL teachers were open to change their beliefs, and thus were more likely to have positive attitudes towards using spontaneous assessments than experienced teachers.

One explanation for novice teachers' strong beliefs about the value of spontaneous assessments is the effect of pre-service teacher programs. Typically, novice teachers will have recently completed a pre-service teacher program, during which they will have learned to make effective instructional decisions, such as using spontaneous assessment opportunities. They would be expected to apply theories and practices from their

preparation program in their teaching and, therefore, likely to show a preference for using spontaneous assessment opportunities.

Second, novice teachers expressed stronger beliefs in the value of formal assessment tasks than experienced teachers. Novice teachers' strong beliefs in the importance of formal assessment tasks might be explained by their previous personal learning experience. The Chinese educational system is characterized by an examination-driven environment (Yu & Jin, 2014), in which formal assessment tasks are used frequently for examination preparation and score improvement. It is possible that novice teachers in the present study had experienced a variety of formal assessments during their schooling, such as exams, numerous textbook exercises, assignments, and recitation tasks. By the time they were recruited as English teachers at primary schools, due to their limited teaching experience, they may have had a preference for the formal assessment methods they had personally experienced.

In contrast, the finding that experienced teachers said they were less likely to use formal assessment tasks might be the result of their classroom teaching experience. Through reflecting on their teaching practice, experienced teachers may have gained a better understanding of students' characteristics as well as the strengths of various assessment methods. They were likely to be aware of the need to select appropriate assessment methods for students, rather than rely on formal assessment tasks.

6.3.2 Effect of previous education in assessment

The present study shows that previous education in assessment did not exert a significant impact on teachers' conceptions of classroom-based assessment purposes or processes. In other words, those teachers who had received training in assessment or attained university coursework on assessment did not put a higher value on classroom-based

assessment. Brown's (2004, 2008) research suggested that New Zealand primary school teachers' conceptions of assessment were independent of assessment training. The finding also lends support to the literature on teacher cognition, which has suggested that teacher training or education programs may not be a vital source of teachers' cognitions (Bao et al., 2016; Junqueira & Kim, 2013; Peacock, 2001; Rahimi & Zhang, 2015; Sun, 2017).

On the whole, the findings do not support the value of assessment training in informing teachers' conceptions of classroom assessment. This may be because the training that the teachers received in bachelor or master level programs, or in in-service teacher education programs, was ineffective. According to Xu and Liu (2009), in China most bachelor's or master's degree programs that prepare teachers of English focus mainly on practical teaching, with little attention given to assessment. Although some master's degree programs offer a course on testing, they are unlikely to include formative assessment. Leung (2004) also points out that formative assessment is not a major component of in-service language teacher professional development programs. It is thus understandable that previous education in assessment did not influence the teachers' conceptions of classroom-based assessment.

6.3.3 Effect of school type

The findings suggest a significant difference between public school and private school teachers' conceptions of formal assessment tasks. While private school teachers agreed on the importance of using formal assessment tasks, public school teachers placed a low value on such methods. Other studies have reported similarly that the type of school in which teachers teach is an important source of their conceptions (Hallam & Ireson, 2003; Zhang & Liu, 2014). Zhang and Liu (2014), for instance, found that teachers

working in schools, with an innovative culture, supportive school authorities, qualified teachers, and high achieving students, were likely to hold progressive beliefs about collaborative learning and teacher-student interaction.

An explanation of private school teachers' favorable conceptions about formal assessment tasks is the distinctive work environment of private schools. As noted by Schulte (2017), the educational infrastructure, teaching resources, and teacher quality of private schools in China are generally not comparable with that of public schools. In China, the majority of primary schools are public schools that are run and funded by local governments at different levels, such as municipal, district, and township public schools. Such schools are well resourced and prestigious to provide all children with equal opportunities for 9-year compulsory education (Brown & Gao, 2015). Private schools have been introduced because of the increasing number of primary school students. To ensure that the quality of teaching in private schools can be recognized by parents and the public, students' achievement in formal assessments is of utmost importance. Under such circumstances, it is perhaps not surprising that teachers in private schools favor formal assessments to a large extent.

The finding that teachers in private schools held strong beliefs in formal assessment methods could also be attributed to their lack of expertise in assessment; in China, private schools are not generally comparable to public schools in regard to teaching resources and government support. It is possible that teachers in private schools have received inadequate professional training and have little expertise or experience in classroom-based assessment. According to Gu (2014) and Inbar-Lourie and Donitsa-Schmidt (2009), teachers rely heavily on traditional assessment methods when lacking expertise in alternative assessment. Hence, it explains these teachers' favorable attitudes towards formal and traditional assessment tasks. In agreement with Inbar-Lourie and

Donitsa-Schmidt (2009), this finding suggests that teachers in private schools need to be provided with professional training in assessment.

6.3.4 Effect of grade level

Teaching at lower grade levels (Grade 1 to Grade 4) and higher grade levels (Grade 5 to Grade 6) did not appear to influence teachers' conceptions of classroom-based assessment. This finding is in contrast with previous studies on teachers' conceptions of assessment in general education contexts, in which grade level was reported to influence teachers' conceptions of classroom-based assessment (Adams & Hsu, 1998; Zhang & Burry-Stock, 2003). In Adams and Hsu's (1998) study, primary school mathematics teachers teaching at higher grade levels (Grade 3 and Grade 4) relied more on objective assessment techniques than those teaching at lower grade levels (Grade 1 and Grade 2). Similarly, Zhang and Burry-Stock's (2003) study showed that as grade level increased, teachers tended to rely heavily on objective formal assessment techniques.

Taken together, the findings of the present study show that teaching experience and school type impacted on teachers' conceptions of classroom-based assessment. From a theoretical perspective, the findings correspond to Borg's (2003) schematic model of teacher cognition, which emphasizes that teachers' classroom teaching experience and contextual factors are two important sources of teachers' conceptions.

6.4 Teachers' classroom-based assessment practices

The third research question examined teachers' practices of classroom-based assessment practices in terms of four dimensions (planning assessment, collecting learning evidence, making professional judgments and providing feedback). In this section, the main features of teachers' CBA practices are discussed, integrating quantitative and qualitative findings.

6.4.1 Planning assessment

While the findings of the survey study show that the teachers self-reported a frequent use of assessment planning activities, the findings of the case study suggest that they did not implement these activities fully. The discrepancy between teachers' self-reported and actual assessment practices demonstrates the limitation of questionnaires in accurately capturing teachers' behaviors. As Dörnyei and Taguchi (2010) point out, the evidence provided by questionnaires is theoretical and thus needs to be interpreted with caution. The discrepancy also provides evidence to support the argument that questionnaires should be complemented with other methods such as classroom observation to get an insight into teachers' actual classroom practices (Borg, 2015; Gao, Zhang, & Tesar, 2020).

Overall, the findings of the present study suggest three key features of primary school EFL teachers' practices of planning assessment. First, consistent with Gu's (2014) study, this study suggests that the instructional objectives were not guided by the curriculum standards. The NECS specifies the overall objectives for primary school English language education and clearly articulates the need for teachers to establish their own instructional objectives according to the standards. The findings show, however, that the curriculum standards were used only when the teachers needed to carefully develop a teaching plan to meet the requirements of school administrators. In the local classrooms, the teachers did not follow the well-designed instructional objectives; they became a 'rhetorical' concept.

The weak alignment between teachers' assessment practices and the curriculum standards may be the result of the top-down policy directive. As claimed by Chen et al. (2014), the central level of policy development in China has little consideration for the

contextual and local differences. Kathy in the case study also complained the national curriculum standards are appropriate for students in big cities such as Beijing and Shanghai rather than students from Chongqing. The teachers were, therefore, less likely to carry out their instruction according to the curriculum standards. It implies that teachers, supervisors, and heads of departments need opportunities to provide feedback to policy makers on how the curriculum standards fit local needs (Waters & Vilches, 2001).

Second, the study suggests that engaging students in summarizing what they had learned at the end of a lesson did not result in a clear understanding of the learning objectives. This finding appears to substantiate the claim made by Turner and Purpura (2016) that poorly clarified learning objectives often result in language learners' cognitive confusion. As there are many instructional opportunities offered in one class, communicating the learning objectives to students is imperative.

Third, the study shows that the teachers did not explain explicitly to students the learning objectives they were expected to achieve in a lesson or a particular language activity, nor did they explain what constituted the criteria of excellent performance. Similar results have been found in other research in L2 contexts (Mui So & Hoi Lee, 2011; Zhou & Deneen, 2016), where teachers tend not to make learning intentions and assessment criteria explicit to students.

Finally, the EFL teachers sometimes considered students' language proficiency and learning needs when selecting assessment methods, a finding similar to that reported by Cheng et al. (2008). This finding suggests that primary school EFL teachers in China attach importance to selecting appropriate assessment methods for effective classroom-based assessment (Chappuis et al., 2012). The study, however, also shows that the

teachers' practices of selecting assessment methods were driven directly by the school's assessment policy, adding weight to the claim that school policy remains a significant factor in teachers' classroom-based assessment practices (Lee et al., 2016).

6.4.2 Collecting evidence of students' learning

The findings of the survey study show that the teachers reported using multiple assessment methods, including spontaneous assessments, student-involving assessment opportunities, and more formal assessments. The case study confirms that the teachers used a wide range of assessments to collect students' learning evidence, a finding that has been commonly reported in previous studies (Chan, 2008; Gan et al., 2018; Gattullo, 2000; Inbar-Lourie & Donitsa-Schmidt, 2009; Yang, 2012).

An explanation for teachers' use of multiple assessment methods could be the characteristics of L2 learners' language competence. According to Shohamy (1994), L2 students' language competence is a complex construct; it cannot be measured entirely by traditional and formal assessments such as tests. Multiple assessment methods have been advocated as having the potential to cater to the varied learning needs of language learners and to encourage language learning (Leung, 2005; Turner & Purpura, 2016). Therefore, it is not surprising that primary school EFL teachers in the current study reported using multiple assessment methods to collect learning evidence.

While the teachers tended to use multiple assessment methods, the findings from both the survey study and the case study indicate that they used these assessment methods with varying degrees of frequency. First, the teachers frequently used spontaneous assessment opportunities (e.g., informal questioning and observation) to modify instruction and provide immediate feedback to students. This finding is

consistent with that of previous studies examining L2 teachers' CBA practices (e.g., Saito & Inoi, 2017; Yang, 2012).

Second, the majority of teachers self-reported that they frequently used formal assessment tasks, such as classroom tests, assignments, textbook exercises, oral reading, and recitation assessment tasks. The findings of the case study confirm the frequent use of formal assessments and suggest that such assessments mainly served a summative function to measure students' learning achievement. Similar findings have been reported in previous studies (Cheng et al., 2008; Guadu & Boersma, 2018; Muñoz et al., 2012; Öz, 2014; Saito & Inoi, 2017; Shohamy et al., 2008).

Teachers' frequent use of formal assessment tasks to serve a summative function maybe because of the large class sizes. Large classes are a marked characteristic of primary school English language education in China. Teachers in such circumstances are likely to use formal assessment tasks (e.g., tests and textbook exercises) as they are considered a convenient and efficient assessment method for teachers to use in such large size classrooms (Cheng & Sun, 2015).

The frequent use of formal assessment might also be explained by the school's assessment policy. The study shows that the teachers employed formal assessment tasks to prepare students for the final assessment conducted by the school administrator. The finding that the policy in the workplace affected teachers' practices is broadly in line with previous research (Basturkmen, 2012; Fang, 1996; Lee et al., 2016).

Another explanation for the frequent use of formal assessments is the examination-oriented culture in the Chinese EFL context. English language education in China is characterized by an examination-oriented culture in which teachers focus mostly on 'teaching to tests' and give priority to learning outcomes (Chen & May, 2016;

Fan, 2018). Driven by this culture, teachers are likely to use traditional and formal assessment tasks. This was especially true for Kathy, a teacher for Grade 6, who expressed that teacher-constructed exercises were used mainly to get students prepared for examination-oriented education in junior secondary schools.

Third, the finding, from both the survey study and case study, that student-involving assessment opportunities, such as self-assessment, peer-assessment, and oral production activities, were the least frequently used by the teachers has also been reported in previous studies (Brumen et al., 2009; Chan, 2008; Chen et al., 2014; Gan et al., 2018; Guadu & Boersma, 2018; Saito & Inoi, 2017; Yang, 2012). The less frequent use of student-involving assessment opportunities could be attributed to teachers' heavy workload. The findings of the case study indicate that Chinese primary school EFL teachers had to fit an overcrowded syllabus, prescribed by their schools, into about 14 lessons each week. Such a tight curriculum schedule left teachers little time to implement assessments involving students for formative purposes regularly. The finding that teachers' workload hindered the implementation of classroom assessment practices has been seen in previous studies conducted in L2 contexts (Büyükkarcı, 2014; Chan, 2008; Mak & Lee, 2014).

Another possible explanation for the less frequent use of student-involving assessments is the lack of training in formative classroom-based assessment. While the survey study shows that the majority of the teachers had received training or education in assessment, the case study suggests that the assessment courses teachers received during master degree's programs focused on testing rather than formative classroom assessment. Teachers also reported receiving insubstantial training in formative assessment during in-service teacher education programs. As a result of a lack of

professional training, it is unsurprising that the teachers in this study seldom employed student-involving assessment opportunities to support learning.

6.4.3 Making professional judgments

Although the finding from the survey study suggests that specific criteria and students' own previous performance were a common reference against which current learning evidence was compared, the case study shows that pupil-referenced assessment and criterion-referenced assessment were used by the teachers relatively infrequently, while the norm-referenced assessment was used most frequently. The mismatch between self-reported and actual practices indicate that questionnaires alone may fail to reveal teachers' real classroom behaviors (Borg, 2015).

Overall, the findings demonstrate that teachers' practices of making judgments did not seem to support student learning to a large extent as there was a lack of criterion-referenced and pupil-referenced assessment. According to McMillan (2008), criterion-referenced assessment has a facilitative role in diagnosing learning and identifying the actions that need to be taken to close learning gaps. The pupil-referenced assessment offers students opportunities to take responsibility for their own learning, enabling them to become self-regulated learners (Harlen & James, 1997; Jacobs & Renandya, 2019). The findings of the present study indicate, however, that the teachers did not have a clear idea of specific criteria that they would apply in class, and did not engage students frequently in pupil-referenced assessments. Consequently, teachers' practices of making judgments might fail to facilitate students' learning.

Moreover, the teachers relied heavily on norm-referenced assessment. Although this type of assessment aimed to engage students in imitating the successful performance of peers, it can foster competition and discourage students (Darandari & Murphy, 2013;

Lok et al., 2016). The norm-referenced assessment might also fail to identify the specific nature of students' performance, offering little help in closing learning gaps (Harlen & James, 1997; McMillan, 2008).

The examination-driven culture in Chinese schools is a possible explanation for the teachers' reliance on norm-referenced assessment; the Chinese educational system is characterized by a culture of standardized testing and accountability (G. Yu & Jin, 2014). In this educational context, in which good academic results play a pivotal role in the lives of students, an emphasis on high academic achievement is likely to draw teachers' attention to norm-referenced assessment. As a result, students are motivated to focus on scores rather than detailed teacher feedback.

6.4.4 Providing appropriate feedback

In the survey study, the teachers reported frequently providing descriptive feedback to identify students' strengths and difficulties in learning and to help students improve their learning. The findings of the case study, however, indicate that the teachers did not provide detailed and descriptive feedback frequently; they provided mainly evaluative feedback such as rewarding, punishing, simple approval, and disapproval comments. This finding of primary school EFL teachers' frequent provision of evaluative feedback is consistent with previous studies (Brumen et al., 2009; Cheng et al., 2004; Lee, 2007a; Zhou & Deneen, 2016).

There are several reasons for teachers' reliance on evaluative feedback, one being the large class sizes in China, which is a distinguishing feature of primary school English language education (Wang, 2009). The classes in the case study were also very large, with around 40 to 50 students in each class. In this teaching context, teachers were unlikely to be able to provide detailed and specific feedback to individual students.

In addition, the findings of the case study point to the tight curriculum workload, which had an influential impact on teachers' feedback practices. In other words, the squeezed teaching syllabus and the time constraints made it hard for the teachers to provide descriptive feedback to students.

Moreover, teachers' feedback practices might be influenced by the lack of professional training in classroom-based assessment. According to Mak and Burn (2016), professional training is crucial to the effectiveness of teachers' feedback practices. As discussed above, while the majority of the EFL teachers had received training or education in assessment, formative classroom assessment was not a major focus. This might explain why primary school EFL teachers in the current study did not fully utilize descriptive feedback to improve young learners' learning.

6.4.5 Teachers' and students' roles in classroom-based assessment

The findings of the current study indicate that primary school EFL teachers' classroom-based assessment practices were "mainly teacher directed and controlled" (Carless, 2011, p. 105), whereas students' role was underplayed in the whole assessment process. Students did not have a clear idea of where they were going, and they were not actively engaged in strategies to close their learning gaps. Similar outcomes have been reported in previous studies on EFL teachers' classroom assessment practices in the Chinese context (Chen, 2009; Chen et al., 2014). For instance, Chen et al. (2014) found that Chinese university EFL teachers played a dominant role in the classroom assessment process and seldom engaged students in self-regulating their own learning.

The traditional grammar-translation approach to language teaching in China is a possible reason for teacher-dominated assessment practices. For many years, the grammar-translation approach has dominated English language teaching in China, by

which Chinese EFL teachers are likely to explain language points and grammar rules through rote memorization and recitation (Wen, 2012). Most classroom interaction is from teachers to students, with little learner agency (Yang & Dai, 2011). This traditional teaching approach might explain why primary school EFL teachers play a central role in classroom-based assessment.

The Chinese educational system, characterized by a traditional teacher-centered classroom environment, in which teachers are considered the knowledge-matter authority and students learn best by listening to teachers, also leads to teacher-dominated assessment practices (S. Yu, Lee, & Mak, 2016). Under such classroom circumstances, it is difficult for primary school EFL teachers to provide opportunities for students to be actively engaged in assessment activities.

To sum up, the findings of primary school EFL teachers' classroom-based assessment practices are consistent with Gu's (2014) and Brumen et al.'s (2009) studies. Such findings indicate that the potential of classroom-based assessment practices in improving teaching and learning in L2 contexts had not been realized. The results thus lend support to a worldwide concern that teachers' practices of classroom-based assessment remain relatively weak (William, 2010b).

6.5 Relationships between teachers' conceptions and practices regarding classroom-based assessment

The findings of the present study suggest that the relationships between primary school EFL teachers' conceptions and practices regarding classroom-based assessment were complex. On the one hand, the teachers' conceptions of classroom-based assessment had an impact on their assessment practices. Data from the survey study indicate that the teachers' conceptions of the purposes of classroom assessment influenced their

assessment practices. That is, teachers with self-reported strong beliefs in the purpose of assessment for learning improvement were more likely to conduct formative assessment practices.

The teachers' conceptions of CBA processes also predicted their assessment practices. For example, their positive beliefs about planning assessment were related to their implementation of assessment planning activities. Teachers who held positive attitudes about student-involving assessment opportunities were more likely to employ assessments such as self-and peer assessment. Likewise, teachers' strong beliefs in the value of formal assessment tasks were closely related to their practices of formal assessments. Teachers who were positively disposed towards making professional judgments, such as using criterion-referenced assessment and pupil-referenced assessment, tended to employ such practices more frequently. Furthermore, teachers' strong beliefs in the value of descriptive feedback seemed to influence their feedback practices.

The findings of the case study also confirm that teachers' conceptions of classroom-based assessment played an important role in their assessment practices. Their beliefs about the value of informal questioning and observation in improving learning appeared to account for their use of these assessment strategies. Their positive attitudes towards self-and peer assessment were evident in their attempts to engage students in evaluating their own work.

These findings, together, suggest that the teachers' conceptions of classroom-based assessment influence how they implement their assessment practices. This provides some support for previous studies (Chan, 2008; Dixon, Hawe, & Parr, 2011; Inbar-Lourie & Donitsa-Schmidt, 2009; Postareff et al., 2012; Rogers et al., 2007;

Young & Jackman, 2014), which have highlighted the significant role that teachers' beliefs about assessment play in shaping their assessment practices. Chan (2008), for example, found that the strong beliefs that teachers held about multiple assessment methods influenced their use of assessment methods to collect learning evidence. Dixon et al. (2011) reported that the teachers' beliefs about the roles of teachers and learners influenced how they provided feedback.

On the other hand, there were discrepancies between primary school EFL teachers' conceptions of classroom-based assessment and their assessment practices. In particular, there was a clear discrepancy between teachers' strong beliefs about the value of planning assessment and the fact that they had not fully utilized planning assessment activities to benefit students' learning. The instructional objectives were not really reflective of the curriculum standards, and the learning objectives and success criteria were not communicated to students. In addition, there was a gap between teachers' stated conceptions and actual practices in regard to making professional judgments. They held strong beliefs about the value of making judgments about learning progress against the prescribed standards (criterion-referenced assessment) and with students' previous performance (pupil-referenced assessment); however, in practice, they conducted norm-referenced assessments frequently. Furthermore, although the teachers attached great importance to descriptive feedback, they actually provided mainly evaluative feedback to enhance students' extrinsic learning motivation.

In summary, the above findings provide evidence that while the teachers said they placed a high value on classroom-based assessment, they did not implement classroom-based assessment frequently and effectively. Similar results have been reported in previous research (Büyükkarcı, 2014; Chan, 2008; Guadu & Boersma, 2018; Muñoz et al., 2012; Rogers et al., 2007; Shohamy et al., 2008). All these studies have suggested

that there is little congruence between L2 teachers' conceptions of classroom-based assessment and their actual assessment practices.

The divergence between primary school EFL teachers' conceptions and practices regarding classroom-based assessment can be explained by a range of factors related to teachers, students, and contexts.

6.5.1 Teacher-related factors

An important teacher-related factor, which may account for the discrepancy between teachers' CBA conceptions and practices, is the lack of expertise in classroom-based assessment. The findings of the case study suggest that the assessment courses that teachers attended during bachelor's or master degree's programs focused mainly on summative assessment of learning, and not assessment for learning. Furthermore, in-service teacher education programs provided insufficient training for teachers to develop a holistic understanding of classroom-based assessment, and to equip teachers with the relevant skills to implement classroom-based assessment effectively. The lack of formal training classroom-based assessment for EFL teachers has also been reported in previous research (Gu, 2014; Lee et al., 2016; Teasdale & Leung, 2000; Xu & Liu, 2009). Consequently, the teachers were unable to implement classroom-based assessment effectively even if they were highly positive towards various classroom-based assessment processes. For example, although the teachers expressed positive beliefs about the benefits of student-involving assessment opportunities, such as self-and peer assessment, they may have been reluctant to implement their beliefs because they were unfamiliar with such assessment methods. Similarly, while the teachers emphasized the importance of descriptive feedback, they tended to provide evaluate feedback as they did not have the necessary skills for providing descriptive feedback.

The suggestion that the discrepancy between teachers' conceptions of classroom-based assessment and their actual assessment practices might be connected to the lack of expertise in classroom-based assessment is consistent with previous studies (Al-Nouh et al., 2014; Brumen & Cagran, 2011; Postareff et al., 2012; Shohamy et al., 2008; Troudi et al., 2009; Young & Jackman, 2014). Shohamy et al. (2008), for example, attributed the gap between teachers' conceptions of assessment methods and their assessment practices to the fact that the teachers were not provided with professional development opportunities to enable them to be assessment literate. Postareff et al. (2012) posited that the discrepancy between teachers' conceptions about the purposes of assessment and their assessment practices could be explained by inadequate training and support received by the teachers.

Another teacher-related factor that possibly contributes to the gap between teachers' CBA conceptions and practices is a hierarchy of beliefs. According to Phipps and Borg (2009), teachers' cognitive systems comprise both peripheral beliefs and core beliefs. Peripheral beliefs "reflect teachers' theoretical or idealistic beliefs—beliefs about what should be—and maybe informed by technical or propositional knowledge" (Phipps & Borg, 2009, p. 382), and thus are subject to change. Teachers' classroom behaviors, on the other hand, are more often powerfully influenced by core beliefs that are more stable and resistant to change (Brownlee, Boulton-Lewis, & Purdie, 2002).

The conceptions reported in the questionnaires, by the teachers in this study, might reflect their peripheral beliefs about CBA theory. Although they agreed with the formative purpose of CBA and said they valued the various components of CBA, they were unlikely to put these conceptions into practice as these beliefs may have been in conflict with more stable core beliefs. For example, the teachers agreed that descriptive feedback should be provided to identify problems in students' learning and to help

students move forward. Nevertheless, they did not put these conceptions into practice but relied heavily on the provision of evaluative feedback. This might be due to their core beliefs about the importance of scores in enhancing students' motivation. Kathy, for instance, claimed that rewards and scores motivated students to higher achievement.

In another example, the teachers in the survey strongly agreed that students should be actively involved in the assessment process through: sharing the learning objectives; carrying out self-and peer assessment; and by teachers providing descriptive feedback to enable students to become self-regulated learners. These theoretical beliefs, however, seemed to contradict more stable beliefs about learning, such as focusing on rote memorization. In the case study, for instance, Amy was observed to show a preference for recitation tasks as she believed that it could help students to memorize what they had learned.

6.5.2 Student-related factors

The second set of factors that may account for the discrepancy between teachers' classroom-based assessment conceptions and their assessment practices is associated with students. First, the gap between teachers' conceptions and practices might be due to students' learning attitudes, needs, and preferences. This study suggests that although teachers held strong beliefs about the value of various classroom assessment processes, they tended to modify their assessment practices according to students' specific needs. For example, Kathy realized that as her students had difficulty in understanding the objectives at the beginning of a lesson, she got them to make a summary of what they had learned at the end of the lesson. Likewise, Doris asked students to summarize what they had learned when she had recognized students were not interested in knowing the objectives at the beginning of the lesson. Knowing that lower achieving students usually

had negative attitudes towards knowing success criteria, Kathy and Doris tended not to set, and share with students, specific assessment criteria. McMillian (2003), similarly, claimed that students' attitudes often create demands, which can affect teachers' assessment practices that are in conflict with what teachers believe about assessment.

The finding that students' attitudes had an impact on teachers' practices is consistent with the findings from other research in education (Nishino, 2012; Phipps & Borg, 2009). For example, Phipps and Borg (2009) reported that students' expectations seemed to outweigh teachers' beliefs about grammar teaching and affected their teaching practices. Nishino (2012) also found that students' preferences and motivation affected the teachers' use of communicative activities. For instance, low motivation among some students made it difficult for teachers to use pair work effectively. The finding of the present study is also consistent with Lee et al.'s (2016) research on EFL teachers' feedback innovation. They reported that students' negative attitudes towards peer feedback could partly explain why teachers did not implement what they had learned from teacher professional development programs fully into their own practices.

Another student-related factor is related to parents' expectations. It is widely acknowledged that in a competitive social context of China, parents' expectations for a high level of academic performance by their children (Lao, 2004) can influence teachers' expectations for students (Rubie-Davies, Peterson, Irving, Widdowson, & Dixon, 2010). Thus, parents' high expectations for academic achievement may account for primary school EFL teachers' reliance on formal assessment tasks and evaluative feedback. Indeed, Amy's case shows that she frequently used oral reading and recitation tasks and reported students' achievements to their parents.

6.5.3 Contextual factors

The third set of factors responsible for the discrepancies between teachers' conceptions and practices regarding classroom-based assessment relates to the contextual challenges that confront teachers. These include large class sizes, heavy curriculum workload, school assessment policies, as well as the examination-driven culture and the Confucian heritage culture.

The first contextual challenge is large class sizes. The study indicates that the classes were very large, with around 40 to 50 students in each class, and an obstacle for the effective implementation of classroom-based assessment. For instance, it was difficult for teachers to provide detailed and descriptive feedback to individual students. Large classes also provided unfavorable conditions to implement student-involving assessment opportunities, such as self-and peer-assessment, as students were not able to be actively engaged in the assessment process. As a result, the teachers preferred to use traditional and formal assessments as they were efficient for large-size classes.

The influence of large class sizes on teachers' CBA practices has been commonly reported in previous studies conducted in L2 contexts (Büyükkarcı, 2014; Chan, 2008; Cheng et al., 2004; Cheng & Sun, 2015; Cheng & Wang, 2007; Guadu & Boersma, 2018; Xu, 2016). Cheng and Wang (2007), in their study about English language teachers' assessment practices in three tertiary institutional contexts (Canada, Hong Kong, and mainland China), found that most Canadian and Hong Kong teachers provided individualized feedback, whereas Chinese teachers did not provide timely and individualized feedback. The major reason given for Chinese teachers' limited uptake of individualized feedback was that they had to manage a large number of students. Büyükkarcı (2014), also, found that most primary school EFL teachers in Turkey were

teaching overcrowded classes (up to 60 students in a class), making it very hard to implement formative assessment. In light of this challenge, and in agreement with Xu and Gary (2018), this study suggests that teachers should be cognitively (assessment literacy) and affectively (commitment to formative assessment) prepared for temporal and contextual conditions to minimize the negative impact of large size classes on classroom-based assessment implementation.

Another contextual challenge is that the teachers had a pressured curriculum and syllabus to teach. Although the teachers appeared to value classroom-based assessment, they had to cover the tight syllabus content and complete the teaching schedule. The teachers, in this study, complained about the need to finish the packed syllabus prescribed in their schools, and that time constraints made it difficult for them to implement self-assessment. They also said that they faced difficulties in providing descriptive and supportive feedback to individual students regularly because of the heavy curriculum workload. In previous research (Chan, 2008; Guadu & Boersma, 2018; Lee et al., 2016; Mak & Lee, 2014), it has been reported that curriculum workload is an important factor influencing teachers' classroom assessment practices. Mak and Lee (2014), for example, found that teachers in Hong Kong had a crowded syllabus, and thus could not implement assessment for learning consistently. Lee et al. (2016), likewise, claimed that the implementation of the feedback became daunting as the teachers had to deal with a tight curriculum.

The tension between primary school EFL teachers' conceptions and practices might also be explained by school assessment policies. Doris and Kathy emphasized the importance of using appropriate assessments to meet students' needs, but they frequently used oral reading tasks for the purpose of rote memorization. Such practice

was influenced by their school's assessment policy, which required students' oral reading performance to be examined by school administrators at the end of the term.

The examination-driven context of China also emerges as significant in influencing teachers' classroom-based assessment practices. As discussed above, the Chinese educational system is characterized by an examination-driven culture (Carless, 2011). For instance, a wide range of standardized tests at school, provincial and national levels are used for the selection of students for admission to primary schools, junior and senior secondary schools, colleges, and universities (Cheng & Curtis, 2007). As a result of this culture, high academic performance is considered as the key to success in life (Cheng & Qi, 2006). In this social culture context, it is possible that the teachers who held positive attitudes towards CBA for learning did not fully utilize CBA practices to promote students' learning. The influence of the examination-driven culture was evident in the case of Doris, where the students were about to enter junior secondary schools. Doris frequently used teacher-constructed exercises and dictation tasks in a summative way to prepare students for examination-oriented education in secondary schools. The finding that high stakes tests exerted a powerful influence on teachers' classroom-based assessment practices is consistent with previous research findings (Chen et al., 2014; Cheng et al., 2008; Cheng & Sun, 2015; Gu, 2014; Öz, 2014; Rogers et al., 2007).

The Confucian heritage culture also poses a substantial challenge to the implementation of classroom-based assessment. It is widely acknowledged that the way of teaching and learning in China has been deeply influenced by the Confucian heritage culture, where there is a heavy emphasis on respect for teachers (Nguyen et al., 2006). As a popular saying indicates, "being a teacher for only one day entitles one to lifelong respect from the student that befits its father" (*yiri weishi zhongshen weifu*) (Hu, 2002). This philosophy contributes to a hierarchical student-teacher relationship, where

teachers are conceptualized as the authoritative figure, and students are regarded as passive recipients (Carless, 2011). Under the influence of this Confucian heritage culture, primary school EFL teachers are likely to control every step of the classroom-based assessment process.

To conclude, the findings of this study suggest that teachers' conceptions of the purposes and processes of classroom-based assessment had an influential role in shaping their assessment practices. There were, however, discrepancies between teachers' conceptions and practices, which were influenced by a variety of factors, including teacher-related factors (e.g., teachers' expertise in classroom-based assessment and the tension between core beliefs and peripheral beliefs), student-related factors (e.g., students' learning needs and parents' expectations), and contextual factors (e.g., class size, curriculum demands, school assessment policies, and examination-driven culture). From a social culture perspective, this study indicates that the implementation of classroom-based assessment is constrained by multiple challenges that stem from key stakeholders (teachers and students) and external school and social contexts.

6.6 Chapter summary

This chapter discussed the major findings of the present study in relation to the existing literature and the context of the study. In conclusion, it is argued that this study advances our understanding of teachers' conceptions and practices regarding classroom-based assessment of young EFL learners.

CHAPTER SEVEN

CONCLUSION, IMPLICATIONS, LIMITATIONS AND RECOMMENDATIONS

7.1 Chapter overview

The final chapter summarizes the major findings and discusses some important implications drawn from this study in relation to theory, research methodology and practice. It concludes with comments on the limitations of this study and recommendations for further research.

7.2 Summary of major findings

The present study investigated primary school EFL teachers' conceptions and practices regarding classroom-based assessment in China. Using a mixed methods research design, this study provides valuable insights into Chinese primary school EFL teachers' conceptions of classroom-based assessment (purposes and processes) and the effects of teacher attributes (teaching experience and previous education in assessment) and work environment (school type and grade level) on teachers' conceptions, and their classroom-based assessment practices. It also contributes to an understanding of the relationships between teachers' CBA conceptions and practices.

First, based on self-reported data, it appears that the teachers were positively disposed to the purpose of using assessment to improve teaching and learning, whereas they slightly agreed with the purpose of making students accountable through assessment and using assessment to prepare students for examinations. Besides, the teachers expressed positive attitudes towards various components of classroom-based assessment. They stated that it was crucial to conduct assessment planning activities,

including establishing clear instructional objectives, sharing the objectives and success criteria with students, and selecting appropriate assessment methods. They agreed that teachers should collect learning evidence by using multiple assessment methods. In reference to specific assessment methods, spontaneous assessment opportunities, and formal assessment tasks were perceived as more important than student-involving assessment opportunities in benefiting students' learning. The teachers showed a preference for criterion-referenced assessment and pupil-referenced assessment and strongly agreed with the use of descriptive feedback.

Second, while previous education in assessment and grade level did not have a significant impact on teachers' conceptions of classroom-based assessment, teaching experience, and school type played an influential role in shaping teachers' conceptions. Specifically, the teachers with little teaching experience (less than 5 years) attached greater importance to both spontaneous assessment opportunities and formal assessment tasks than those with medium teaching experience (6 to 20 years). Teachers from private schools expressed a greater preference for formal assessment tasks than did teachers from public schools.

Third, the teachers had not fully implemented classroom-based assessment practices to support students' learning. As regards planning assessment, the instructional objectives were not guided by the curriculum standards, and the learning objectives and success criteria were not clarified to students. The selection of assessment methods was driven largely by their schools' assessment policy. Although the teachers used multiple assessment methods, the specific methods were used to varying degrees. Spontaneous assessment opportunities and formal assessment tasks were more frequently used, whereas student-involving assessment opportunities were less frequently used. Teachers' practices of making judgments seemed not to be used to support students' learning as

there was a reliance on norm-referenced assessments rather than criterion-referenced and pupil-referenced assessments. The teachers frequently provided evaluative feedback to enhance students' extrinsic motivation, with little use of descriptive feedback.

Finally, the relationships between teachers' conceptions and practices regarding classroom assessment were found to be complex. On the one hand, the teachers' conceptions of classroom-based assessment had an important role in informing their assessment practices. Teachers who had expressed positive attitudes towards the formative purpose and the various components of classroom-based assessment were more likely to employ formative assessment practices. On the other hand, there were discrepancies between teachers' conceptions of classroom-based assessment and their actual assessment practices, particularly in relation to planning assessment, making professional judgments, and providing appropriate feedback. The limited congruence may be attributed to teachers' lack of expertise in classroom-based assessment, students' learning attitudes, parents' expectations, and contextual challenges such as large class sizes, curriculum demands, school assessment policies, examination-driven culture, and the Confucian heritage culture.

7.3 Implications

A number of important implications are drawn from this study, which is discussed with regard to theory, research methodology, and practice.

7.3.1 Theoretical implications

This study has several theoretical implications. First, it illustrates the importance of a comprehensive investigation of the implementation of classroom-based assessment. As the literature review shows, most of the existing studies examining classroom-based assessment in L2 contexts focus on one or two components, such as using multiple

assessment methods and providing feedback. As Leung (2005) stresses, there is a need to explore teachers' CBA practices more comprehensively to further develop grounded research base on CBA. In response to this challenge, the present study has investigated CBA implementation from four dimensions: planning assessment, using multiple assessment methods, making professional judgments, and providing appropriate feedback. It thus contributes to a more sophisticated understanding of the current status of CBA implementation and generates further insights into how CBA can be effectively implemented.

Second, the study provides valuable insights into primary school EFL teachers' conceptions of classroom-based assessment. While there is strong evidence that teachers' conceptions have an extensive role in shaping their practices, there is a limited volume of research on L2 teachers' conceptions of CBA, particularly when compared with the existing body of research on teachers' CBA practices. This study has looked into how primary school EFL teachers conceptualize CBA; it posits that the teachers hold positive beliefs about the formative purpose and key processes of CBA.

Moreover, the study enhances our understanding of the sources of teachers' conceptions. As Borg's (2003) model suggests, the development of teachers' conceptions can be influenced by a number of factors, including prior learning experience, teacher education, teaching experience, and contextual factors. This study has identified the sources of teacher conceptions by investigating how teacher attributes (such as teaching experience and previous education in assessment) and work environments (such as school type and grade level) influence primary school EFL teachers' conceptions of CBA. The findings from the present study argue for the powerful influence of teaching experience and school type on teachers' conceptions.

This study has demonstrated the complex relationships between teachers' conceptions and practices regarding classroom-based assessment. It adds to the existing literature by showing that the conceptions of assessment held by the teachers have an influential role in informing their assessment practices. It also provides evidence that teachers' self-reported conceptions are not always reflected in their classroom practice. A variety of factors that can contribute to the discrepancy between teachers' conceptions and practices regarding CBA have been identified, including a lack of expertise in assessment, students' learning needs, parents' expectations, large class sizes, curriculum demands, school assessment policies, examination-driven culture, and Confucian heritage culture.

Finally, this study makes a valuable contribution to advancing the field of assessment of young language learners. Most of the existing research on CBA in L2 contexts has been conducted with teachers from secondary schools or universities, with little attention paid to young language learners. As Hasselgreen (2012) argues, YLL assessment has a unique research agenda and that there is an urgent need to investigate thoroughly every aspect of this field. This study has identified the important role that teachers' conceptions play in their practices of assessing YLLs as well as a number of factors that influence the way how teachers assess YLLs. Overall, this deepens our understanding of the current status of classroom-based assessment of YLLs.

7.3.2 Methodological implications

In terms of methodological implications, this study proves the value of using a mixed methods research design to explore the relationships between teachers' conceptions and practices regarding classroom-based assessment. It suggests that while the employment of quantitative methods helps explore the general characteristics of teachers'

conceptions and practices, a follow-up case study can provide deeper insights into teachers' classroom practices, as well as the rationales behind the practices. Taken together, the multiple forms of evidence can help develop a more sophisticated understanding of the relationships between teachers' conceptions and practices.

The study also highlights the potential problem of questionnaires in studying teachers' conceptions and practices. As discussed in the previous chapter, there was a discrepancy between teachers' self-reported CBA practices revealed through questionnaires and their actual assessment practices identified through the case study. This discrepancy provides evidence to support the claim about the limitation of questionnaires in the investigation of teachers' conceptions. Borg (2006), for example, in a methodological review of language teacher cognition, argues that questionnaires are inadequate in their ability to capture teachers' real classroom practices. Furthermore, Kane, Sandretto and Heath (2002) argues that "research that examines only what university teachers say about their practices and does not directly observe what they do is a risk of telling half of the story" (p. 177). An important implication is that classroom observations can be used as a complementary instrument to establish a comprehensive understanding of teachers' conceptions and classroom practices.

7.3.3 Practical implications

From a practical perspective, this study provides important implications for policy makers and teacher educators who are interested in managing classroom-based assessment reform. These implications can be of relevance to other EFL contexts that share similar conditions.

7.3.3.1 Implications for policy makers

This study provides insights into the role of teachers' conceptions in influencing their assessment practices, as well as teachers' limited use of formative classroom assessment practices, which have implications for policy makers. First of all, the findings of the current study show that the curriculum development process appeared to take little account of local implementers' needs, which posed impediments to teachers' assessment practices. This suggests that those responsible for implementing curricula reform in the educational system should be invited to contribute to the needs analysis process to share their knowledge, beliefs, and needs with reform planners. As claimed by Waters and Vilches (2001), the need analysis process is crucial in ensuring the reform fits with the prevailing beliefs and preconceptions of the implementers. In the case of curriculum development, consultation meetings should be held for teachers, headteachers, teacher educators, supervisors, and parents to provide feedback to the curriculum designers on how well draft curricula fit with the current situation of teaching and learning. Once the curriculum has been implemented, a nationwide needs analysis can be conducted among various implementers to gain an understanding of their attitudes towards and challenges experienced during the curriculum implementation. Responses from those responsible for implementation can help policy planners improve the curricula.

Second, the findings of the current study show that Chinese primary school EFL teachers lacked experience and expertise in classroom-based assessment, creating a great challenge for effective CBA implementation. An important implication of such is that teachers must be given opportunities for professional development so that they can be equipped with the essential knowledge and skills required for effective classroom-based assessment. To achieve this, school leaders should encourage teachers to attend relevant professional development programs, workshops, conferences, and seminars to

engage them in ongoing professional learning to advance their knowledge and skills of classroom-based assessment. As noted by Clarke and Hollingsworth (2002), ongoing and life-long professional learning is more effective than one-shot programs in teacher change.

Third, it is important to create professional learning communities in schools. The findings of this study show that while most teachers held strong and positive attitudes towards classroom-based assessment, they did not adopt formative assessment practices consistently and regularly. This may reflect that primary school EFL teachers tend to implement assessment practices on an individual level, without collaborative advice and support from community members. As suggested by Stoll, Bolam, McMahon, Wallace, and Thomas (2006), professional learning communities emphasize collaborative efforts rather than individual autonomy and thus have “considerable promise for capability building for sustainable improvement” (p. 221). They contribute to the shared values and vision among the entire school (Dufour, 2006), which enables schools to have a solid foundation for moving forward to implement educational reform (Hipp & Huffman, 2007). Professional communities can also create a collaborative learning culture in which staff works together to analyze and solve problems and improve their classroom practices (Dufour, 2006). A supportive professional learning community, therefore, can play a vital role in the successful implementation of classroom-based assessment. To foster a supportive professional community, meetings and discussions can be held to convey the principles and purposes of classroom-based assessment to staff at all levels of the school. Whole school communities can share their experiences in adopting assessment practices and in navigating the challenges that emerge in the implementation process.

Finally, to enable the teachers to fully implement classroom-based assessment, the support from school administrators is needed. The findings of the current study suggest that a heavy curriculum workload constrained the implementation of classroom-based assessment. This suggests that teachers should have greater autonomy in planning their teaching schedule so that they can put new assessment ideas into practice, as well as reflect on their practices. As Lee et al. (2016) point out, teachers' flexibility in adjusting the curriculum is imperative for innovative initiatives.

7.3.3.2 Implications for teacher educators

The finding of a lack of teachers' experience and expertise in classroom assessment emphasizes the importance of including classroom-based assessment in teacher education programs. As suggested by researchers (Giraldo, 2018; Vogt & Tsagari, 2014), language teachers should be equipped with not only a systematic knowledge of language pedagogy but also skills and knowledge in teacher-based assessment. It is, thus, recommended that teacher education programs are designed to include content knowledge about classroom-based assessment and to offer teachers practical guidelines on how to incorporate theory into practice in the classroom. Imparting the knowledge of CBA alone, however, is inadequate. To ensure the successful implementation of CBA, it is also important to raise awareness among teachers of the purposes and key principles of classroom-based assessment (Wang, 2020). As the teachers' conceptions in this study were shown to influence the way they implemented assessment practices, attention needs to be given to changing or developing teachers' conceptions of classroom-based assessment. An implication for teacher education is that teacher educators should equip teachers with the necessary skills to reflect upon their own assessment practices (Howard, 2003) so that they become aware of the purposes of CBA and the influence that CBA can have on teaching and learning.

In addition, it is essential that teacher educators help raise pre-service or in-service teachers' awareness of potential impediments that will constrain the full implementation of classroom-based assessment. As argued by Golombek and Johnson (2004), teachers' recognition of challenges in the classroom context is "a driving factor in teacher development" (p. 324). The discrepancy between teachers' CBA conceptions and their actual practices found in this study indicates that contextual factors, such as large class sizes, heavy curriculum load, and exam pressure, are major challenges to the effective implementation of classroom-based assessment. In this sense, it is suggested that teacher educators organize discussions on possible impediments to the successful implementation of CBA in the teaching context.

7.5 Limitations of the study

Despite the implications drawn from this study, several limitations need to be addressed. First, the research question concerning teachers' conceptions of classroom-based assessment was addressed by the survey study alone. While questionnaires are efficient in generating a general pattern of teachers' conceptions of CBA, they offer limited insights into teachers' conceptions. As Borg (2006) notes, teachers' responses to questionnaires may fail to reflect the full range of their conceptions as they are limited by the items designed by the researcher. If post-questionnaire interviews had been conducted, it could have been helpful in providing deeper insights into teachers' conceptions of CBA.

Another limitation regarding the survey study is that the questionnaire had not been subject to large-scale piloting, which could have an impact on its validity. Chapter Four outlined how the questionnaire was developed. While there was careful attention to item pool generation, content validity, translation, and piloting, due to time

constraints, only a small scale piloting was conducted to obtain feedback from potential participants. Therefore, further modification and large-scale piloting would have improved the reliability and validity of the questionnaire.

Third, while the case study was valuable in exploring in-depth teachers' classroom assessment practices and the underlying rationales, a limited amount of qualitative data was gathered. Only three cases were studied due to time and research budget constraints, with all the three teachers recruited from one province. Although the findings of the case study have reflected the unique features of the three individual teachers, caution should be exercised when considering the relevance of the findings to teachers in other contexts.

A fourth limitation is that the second research question regarding the effects of teacher attributes and work environments on teachers' conceptions has not addressed teachers' prior learning experience. As Borg's (2003) model of teacher cognition shows, prior learning experience, teaching experience, teacher education, and contextual factors are four major sources of teachers' cognition. Examining the effect of teachers' learning experience would have contributed to a full exploration of the sources of teachers' CBA conceptions.

Finally, the study only investigated the conceptions and practices regarding CBA of primary school EFL teachers. It did not include any other stakeholders in the educational system, such as students, headteachers, teacher educators and policy makers. Exploring the conceptions of different education stakeholders about CBA and how their conceptions are implemented in practice could have provided further insights into how to promote effective CBA implementation.

7.6 Recommendations for further research

This study has demonstrated that research which examines the general characteristics of teachers' conceptions and practices regarding classroom-based assessment through a survey study, and which further explores teachers' actual practices and underlying rationales through a case study, can contribute to understanding the complex relationships between teachers' conceptions and practices. Building on the findings of this study, recommendations for further research on classroom-based assessment and teacher cognition are suggested.

First, this study has investigated EFL teachers' classroom-based assessment of young learners in China. Further research of this kind in other education contexts is imperative as the assessment of young language learners is a newly emerging field that warrants further research attention. Results, either similar or different, will contribute to a more sophisticated picture of the implementation of classroom-based assessment of young language learners.

Second, this study has examined the effects of teacher attributes (teaching experience and previous education in assessment) and work environments (school type and grade level) on teachers' conceptions of CBA. Further work on the effect of teachers' prior learning experience, perhaps using different research methods, will lead to a full understanding of the sources of teachers' conceptions of CBA.

Finally, it is recommended that further studies are carried out with different education stakeholders working in similar contexts. The participants in the current study were primary school EFL teachers; it will be interesting to investigate how students, headteachers, and teacher educators perceive and implement classroom-based assessment in the same context.

7.7 Conclusion

To conclude, this study aims to enhance our understanding of classroom-based assessment of young language learners from the perspective of teacher cognition. It shows that Chinese primary school EFL teachers hold positive beliefs about the purpose of using assessment to improve teaching and learning, as well as strong beliefs about the value of the key processes of CBA in improving learning. It also confirms that teaching experience and school type play an important role in shaping teachers' conceptions of classroom-based assessment. The study provides evidence that the uptake of formative classroom-based assessment practices by primary school EFL teachers in China is limited. The most influential contribution made by the present study lies in the insights it provides into the complex relationships between teachers' conceptions and practices regarding classroom-based assessment. As well as demonstrating that teachers' conceptions influence their assessment practices, it indicates that teachers' conceptions are not always reflected in their assessment practices; it also identifies factors that may account for the lack of consistency between teachers' conceptions and classroom practices.

REFERENCES

- Abedi, J. (2010). Research and recommendation for formative assessment with English language learners. In H. L. Andrade & G. J. Cizek (Eds.), *The handbook of formative assessment* (pp. 181–197). New York, NY: Routledge.
- Abu-Alhija, F. N. (2007). Large-scale testing: Benefits and pitfalls. *Studies in Educational Evaluation*, 33, 50–68.
- Adams, T. L., & Hsu, J. W. Y. (1998). Classroom assessment: Teachers' conceptions and practices in mathematics. *School Science and Mathematics*, 98(4), 174–180.
- Adams, W. C. (2015). Conducting semi-structured interviews. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *The handbook of practical program evaluation* (3rd ed., pp. 492–505). San Francisco, CA: Wiley.
- Adamson, B. (2004). *China's English*. Hong Kong, China: Hong Kong University Press.
- Airasian, P. W. (2012). *Classroom assessment: Concepts and applications* (7th ed.). New York, NY: McGraw-Hill.
- Airasian, P. W., & Abrams, L. (2003). Classroom student evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (pp. 533–548). Dordrecht, the Netherlands: Springer.
- Akyel, A. (1997). Experienced and student EFL teachers' instructional thoughts and actions. *Canadian Modern Language Review*, 53, 678–704.
- Al-Nouh, N. A., Taqi, H. A., & Abdul-Kareem, M. M. (2014). EFL primary school teachers' attitudes, knowledge and skills in alternative assessment. *International Education Studies*, 7(5), 68–84.
- Alderson, C., & Banerjee, J. (2001). Language testing and assessment (Part I). *Language Teaching*, 34, 213–236.

- Aliakbari, M., & Jamalvandi, B. (2012). Realization of culture in English textbooks in Chinese high school. *Journal of Pan-Pacific Association of Applied Linguistics*, 16(2), 89–101.
- Allal, L. (2016). The co-regulation of student learning in an assessment for learning culture. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 259–273). Heidelberg, Switzerland: Springer.
- Alshenqeti, H. (2014). Interviewing as a data collection method: A critical review. *English Linguistics Research*, 3(1), 39–45.
- Ames, C. (1992). Classroom: goals, structures and student motivation. *Journal of Educational Psychology*, 84(3), 261–271.
- Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. L. Andrade & G. J. Cizek (Eds.), *The handbook of formative assessment* (pp. 90–105). New York, NY: Routledge.
- Andrade, H. L. (2018). Feedback in the context of self-assessment. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 376–408). Cambridge, England: Cambridge University Press.
- Andrade, H. L., Bennett, R. E., & Cizek, G. J. (2019). *The handbook of formative assessment in the disciplines*. New York, NY: Routledge.
- Andrade, H. L., & Brookhart, S. M. (2016). The role of classroom assessment in supporting self-regulated learning. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 293–309). Heidelberg, Switzerland: Springer.
- Andrade, H. L., & Heritage, M. (2018). *Using formative assessment to enhance learning, achievement, and academic self-regulation*. New York, NY: Routledge.

- Andrews, S. (2003). “Just like instant noodles”: L2 teachers and their beliefs about grammar pedagogy. *Teachers and Teaching: Theory and Practice*, 9(4), 351–375.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback—A case-study. *System*, 30, 207–223.
- Araceli Ruiz-Primo, M., & Furtak, E. M. (2006). Exploring teachers’ informal formative assessment practices and students’ understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84.
- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge, England: University of Cambridge School of Education.
- Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Cambridge, England: Cambridge University Faculty of Education.
- Bailey, A. L. (2008). Assessing the language of young learners. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., pp. 379–398). New York, NY: Springer.
- Bailey, A. L. (2017). Theoretical and developmental issues to consider in the assessment of young learners’ English language proficiency. In M. K. Wolf & Y. G. Butler (Eds.), *English language proficiency assessment for young learners* (pp. 24–39). New York, NY: Routledge.
- Bailey, A. L., & Osipova, A. V. (2016). *Children’s multilingual development and education: Fostering linguistic resources in home and school context*. Cambridge, England: Cambridge University Press.
- Baker, L. M. (2006). Observation: A complex research method. *Library Trends*, 55(1), 171–189.
- Bao, C., Zhang, L. J., & Dixon, H. (2016). Teacher narratives as theorisation of teaching: A Chinese teacher’s perspectives on communicative language teaching

- (CLT). *Journal of the Applied Linguistics Association of New Zealand*, 22(2), 35–50.
- Bao, R. (2019). Oral corrective feedback in L2 Chinese classes: Teachers' beliefs versus their practices. *System*, 82, 140–150.
- Barnard, R., & Burns, A. (2012). Introduction. In R. Barnard & A. Burns (Eds.), *Researching language teacher cognition and practice: International case studies* (pp. 1–10). Bristol, England: Multilingual Matters.
- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. G. Gill (Eds.), *The handbook of research on teachers' beliefs* (pp. 229–233). New York : Routledge.
- Basturkmen, H. (2012). Review of research into the correspondence between language teachers' stated beliefs and practices. *System*, 40, 282–295.
- Basturkmen, H., Loewen, S., & Ellis, R. (2004). Teachers' stated beliefs about incidental focus on form and their classroom practices. *Applied Linguistics*, 25(2), 243–272.
- Belotto, M. J. (2018). Data analysis methods for qualitative research: Managing the challenges of coding, interrater reliability, and thematic analysis. *The Qualitative Report*, 23(11), 2622–2633.
- Berry, R. (2011). Assessment reforms around the world. In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp. 89–102). New York, NY: Springer.
- Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *The Sage handbook of mixed methods in social and behavioral research* (2nd ed., pp. 95–118). Thousand Oaks, CA: Sage.

- Black, P. (1998). *Testing: friend or foe? Theory and practice of assessment and testing*. London, England: Falmer.
- Black, P. (2001). Formative assessment and curriculum consequences. In D. Scott (Ed.), *Curriculum and assessment* (pp. 7–23). Westport, CT: Ablex.
- Black, P. (2015). Formative assessment – An optimistic but incomplete vision. *Assessment in Education: Principles, Policy and Practice*, 22(1), 161–177.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, 25(6), 551–575.
- Blaxter, L., Hughes, C., & Tight, M. (2010). *How to research* (4th ed.). Berkshire, England: McGraw-Hill.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *The handbook of formative and summative evaluation of student learning*. New York, NY: McGraw-Hill.
- Borg, S. (2001). Self-perception and practice in teaching grammar. *ELT Journal*, 55(1), 21–29.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81–109.
- Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London, England: Continuum.

- Borg, S. (2011). The impact of in-service teacher education on language teachers' beliefs. *System*, 39, 370–380.
- Borg, S. (2015). *Teacher cognition and language education: Research and practice*. London, England: Bloomsbury.
- Borg, S. (2019). Language teacher cognition: Perspectives and debates. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 1149–1169). Cham, Switzerland: Springer.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18(5), 529–549.
- Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education: Learning for the longer term*. London, England: Taylor and Francis.
- Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist*, 44(3), 576–578.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101.
- Breen, M. P., Hird, B., Milton, M., Oliver, R., & Thwaite, A. (2001). Making sense of language teaching: Teachers' principles and classroom practices. *Applied Linguistics*, 22(4), 470–501.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy and Practice*, 11(1), 7–26.
- Brookhart, S. M. (1997). A framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education*, 10(2), 161–180.

- Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy and Practice*, 8(2), 153–169.
- Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429–458.
- Brookhart, S. M. (2008). *How to give effective feedback to your students*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Brookhart, S. M. (2018). Summative and formative feedback. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 52–78). Cambridge, England: Cambridge University Press.
- Brookhart, S. M., Moss, C., & Long, B. (2008). Formative assessment that empowers. *Educational Leadership*, 66(3), 52–57.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy and Practice*, 11(3), 301–318.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological Reports*, 99, 166–170.
- Brown, G. T. L. (2008). Assessment literacy training and teachers' conceptions of assessment. In C. M. Rubie-Davies & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning* (pp. 285–302). New York, NY: Nova Science.
- Brown, G. T. L. (2013). Student self-assessment. In J. H. McMillian (Ed.), *The Sage handbook of research on classroom assessment* (pp. 367–393). Los Angeles, CA: 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.

- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice*, 16(3), 347–363.
- Brown, G. T. L., Peterson, E. R., & Yao, E. S. (2016). Student conceptions of feedback: Impact on self-regulation, self-efficacy, and academic achievement. *British Journal of Educational Psychology*, 86(4), 606–629.
- Brownlee, J., Boulton-Lewis, G., & Purdie, N. (2002). Core beliefs about knowing and peripheral beliefs about learning: Developing a holistic conceptualisation of epistemological beliefs. *Australian Journal of Educational and Developmental Psychology*, 2, 1–16.
- Brumen, M., & Cagran, B. (2011). Teachers' perspectives and practices in assessing young foreign language learners in three Eastern European countries. *Education 3-13*, 39(5), 541–559.
- Brumen, M., Cagran, B., & Rixon, S. (2009). Comparative assessment of young learners' foreign language competence in three Eastern European countries. *Educational Studies*, 35(3), 269–295.
- Budiman, A. (2017). Behaviorism in foreign language teaching. *English Franca: Academic Journal of English Language and Education*, 1(2), 101–114.
- Buhagiar, M. (2006). The classroom assessment cycle within the alternative assessment paradigm: Exploring the role of the teacher. *Journal of Maltese Education Research*, 4(2), 17–36.
- Burns, A., Freeman, D., & Edwards, E. (2015). Theorizing and studying the language-teaching mind: Mapping research on language teacher cognition. *Modern*

- Language Journal*, 99(3), 585–601.
- Butler, Y. G. (2009). Elementary school students' foreign language performance? A case study from South Korea. *TESOL Quarterly*, 43(3), 417–444.
- Butler, Y. G. (2011). The implementation of communicative and task-based language teaching in the Asia-Pacific region. *Annual Review of Applied Linguistics*, 31, 36–57.
- Butler, Y. G. (2015). Parental factors in children's motivation for learning English: A case in China. *Research Papers in Education*, 30(2), 164–191.
- Butler, Y. G. (2016). Assessing young learners. In D. Tsagari & J. Banerjee (Eds.), *The handbook of second language assessment* (pp. 359–375). Boston, MA: De Gruyter Mouton.
- Butler, Y. G. (2019). Assessment of young English learners in instructional settings. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 477–496). Cham, Switzerland: Springer.
- Butler, Y. G., & Zeng, W. (2014). Young foreign language learner's interactions during task-based paired assessments. *Language Assessment Quarterly*, 11(1), 45–75.
- Büyükkarcı, K. (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction*, 7(1), 107–120.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Routledge.
- Cabaroglu, N., & Roberts, J. (2000). Development in student teachers' pre-existing beliefs during a 1-year PGCE programme. *System*, 28, 387–402.
- Calderhead, J. (1996). Teachers: beliefs and knowledge. In D. Berliner & B. Calfee (Eds.), *The handbook of educational psychology* (pp. 709–725). New York, NY:

Macmillan.

- Cameron, L. (2003). Challenges for ELT from the expansion in teaching children. *ELT Journal*, 57(2), 105–112.
- Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. London, UK: Routledge.
- Carless, D., & Lam, R. (2014). The examined life: Perspectives of lower primary school students in Hong Kong. *Education 3-13*, 42(3), 313–329.
- Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 83(1), 1–6.
- Chan, Y. (2008). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English Language Teaching*, 7(1), 37–62.
- Chappuis, J. (2005). Helping students understand assessment. *Educational Leadership*, 63(3), 39–43.
- Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom assessment for student learning: Doing it right--Using it well* (2nd ed.). Boston, MA: Pearson.
- Chen, Q. (2009). The potential barriers to college English assessment policy change in China: A sociocultural perspective. In B. Garrick, S. Poed, & J. Skinners (Eds.), *Educational planet shapers: Researching, hypothesising, dreaming the future* (pp. 114–126). Brisbane: Post Pressed.
- Chen, Q., & May, L. (2016). Chinese EFL students' response to an assessment policy change. In G. Yu & Y. Jin (Eds.), *Assessing Chinese learners of English: The language constructs, consequences and conundrums* (pp. 199–218). Basingstoke, England: Palgrave.
- Chen, Q., May, L., Klenowski, V., & Kettle, M. (2014). The enactment of formative

- assessment in English language classrooms in two Chinese universities: Teacher and student responses. *Assessment in Education: Principles, Policy and Practice*, 21(3), 271–285.
- Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, 25(1), 15–37.
- Cheng, L. (2015). Introduction to language classroom assessment. In *Language Classroom Assessment* (pp. 1–40).
- Cheng, L., & Curtis, A. (2010). The impact 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. 267–273). New York, NY: Routledge.
- Cheng, L., & Qi, L. (2006). Description and examination of the National Matriculation English Test. *Language Assessment Quarterly*, 3(1), 53–70.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360–389.
- Cheng, L., Rogers, T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment and Evaluation in Higher Education*, 33(1), 9–32.
- Cheng, L., & Sun, Y. (2015). Teachers grading decision making: Multiple influencing factors and methods. *Language Assessment Quarterly*, 12(2), 213–233.
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, 48(4), 436–470.
- Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, 4(1), 85–107.

- Chik, A., & Besser, S. (2011). International language test taking among young learners: A Hong Kong case study. *Language Assessment Quarterly*, 8(1), 73–91.
- Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39–62.
- Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *The handbook of formative assessment* (pp. 3–17). New York, NY: Routledge.
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *The handbook of research on teaching* (3rd ed., pp. 255–296). New York, NY: Macmillan.
- Clark, C. M., & Yinger, R. J. (1977). Research on teacher thinking. *Curriculum Inquiry*, 7(4), 279–304.
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher development. *Teaching and Teacher Education*, 18, 947–967.
- Clarke, S. (2005). *Formative assessment in action: Weaving the elements together*. London, UK: Hodder and Stoughton.
- Clarke, S., Timperley, H., & Hattie, J. (2001). *Unlocking formative assessment: Practical strategies for enhancing students' learning in the primary and intermediate classroom*. London, England: Hodder and Stoughton.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Routledge.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: Sage.
- Couper, G. (2017). Teacher cognition of pronunciation teaching: Teachers' concerns and issues. *TESOL Quarterly*, 51(4), 820–843.

- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101–116.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: Sage.
- Crookes, G., & Arakaki, L. (1999). Teaching ideas and work conditions in an ESL program. *TESOL Journal*, 8(1), 15–19.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481.
- Cumming, A. (2009). Language assessment in education: Tests, curricula, and teaching. *Annual Review of Applied Linguistics*, 29, 90–100.
- Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 3–49). Los Angeles, CA: Evaluation, Dissemination and Assessment Center, California State University.
- Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in Educational Evaluation*, 45, 62–69.
- Darandari, E., & Murphy, A. (2013). Assessment of student learning. In L. Smith (Ed.), *Higher education in Saudi Arabia: Achievement, challenges and opportunities* (pp. 61–71). New York, NY: Springer.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL assessment practices in Australian and Hong Kong secondary schools. *Language*

- Testing*, 21(3), 305–333.
- Davison, C. (2007). Views From the Chalkface: English Language School-Based Assessment in Hong Kong. *Language Assessment Quarterly*, 4(1), 37–68.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415.
- DeMarrais, K. (2004). Qualitative interview studies: Learning through experience. In K. DeMarrais & S. D. Lapan (Eds.), *Foundations for research: Methods of inquiry in education and the social sciences* (pp. 51–68). Mahwah, NJ: Erlbaum.
- Denzin, N. K., & Lincoln, Y. S. (2017). *The Sage handbook of qualitative research* (5th ed.). Los Angeles, CA: Sage.
- Derrick, D. J. (2016). Instrument reporting practices in second language research. *TESOL Quarterly*, 50(1), 132–153.
- Dixon, H. R., Hawe, E., & Parr, J. (2011). Enacting assessment for learning: The beliefs practice nexus. *Assessment in Education: Principles, Policy and Practice*, 18(4), 365–379.
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into Practice*, 55(2), 153–159.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Dörnyei, Z., & Csizér, K. (2012). How to design and analyze surveys in second language acquisition research. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 74–94). Malden, MA: Wiley-Blackwell.
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). New York, NY:

- Routledge.
- Dufour, R. (2006). What is a “ Professional Learning Community” ? *Educational Leadership*, 61(8), 6–11.
- Dweck, C. S. (1996). Social motivation: Goals and social-cognitive processes. In J. Juvonen & K. R. Wentzel (Eds.), *Social motivation: Understanding children’s school adjustment* (pp. 181–195). New York, NY: Cambridge University Press.
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development*. Philadelphia, PA: Psychology Press.
- Elwood, J., & Klenowski, V. (2010). Creating communities of shared practice: The challenges of assessment use in learning and teaching. *Assessment and Evaluation in Higher Education*, 27(3), 243–256.
- Eren, A. (2013). Consonance and dissonance between Turkish prospective teachers’ values and practices: Conceptions about teaching, learning, and assessment. *Australian Journal of Teacher Education*, 35(3), 27–48.
- Esterberg, K. G. (2002). *Qualitative methods in social research*. Boston, MA: McGraw-Hill.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4.
- Falchikov, N. (1995). Improving feedback to and from students. In P. Knight (Ed.), *Assessment for learning in higher education* (pp. 157–166). London, England: Kogan Page.
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. New York, NY: Routledge.

- Fan, J. (2018). A survey of English language testing practices in China: Students' and teachers' perspectives. In D. Xerri & P. V. Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 283–300). New York, NY: Springer.
- Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38*(1), 47–65.
- Farrell, T. S. C., & Ives, J. (2015). Exploring teacher beliefs and classroom practices through reflective practice: A case study. *Language Teaching Research, 19*(5), 594–610.
- Fauziati, E. (2016). *Applied linguistics: Principles of foreign language teaching, learning, and researching*. Surakarta: Era Pustaka Utama.
- Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. *Review of Research in Education, 20*, 3–56.
- Field, A. (2009). *Discovering statistics using IBM SPSS statistics* (3rd ed.). Thousand Oaks, CA: Sage.
- Flick, U. (2018). Triangulation in data collection. In U. Flick (Ed.), *The Sage handbook of qualitative data collection* (pp. 527–544). Thousand Oaks, CA: Sage.
- Fowler, F. J. (2013). *Survey research methods* (5th ed.). Los Angeles, CA: Sage.
- Frederiksen, J. R., & White, B. Y. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.
- Freeman, D. (2002). The hidden side of the work: Teacher knowledge and learning to teach. A perspective from north American educational research on teacher education in English language teaching. *Language Teaching, 35*(01), 1–13.
- Fu, T., & Nassaji, H. (2016). Corrective feedback, learner uptake, and feedback

- perception in a Chinese as a foreign language classroom. *Studies in Second Language Learning and Teaching*, 6(1), 159–181.
- Fung, L., & Chow, L. P. Y. (2002). Congruence of student teachers' pedagogical images and actual classroom practices. *Educational Research*, 44(3), 313–321.
- Gagné, R. M. (1965). *Conditions of learning*. New York, NY: Rinehart and Winston.
- Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy and Practice*, 20(2), 150–169.
- Gan, Z., Leung, C., He, J., & Nang, H. (2018). Classroom assessment practices and learning motivation: A case study of Chinese EFL students. *TESOL Quarterly*, 53(2), 514–529.
- Gao, L. X., Zhang, L. J., & Tesar, M. (2020). Teacher cognition about sources of English as a foreign language (EFL) listening anxiety: A qualitative study. *Linguistic and Philosophical Investigations*, 19, 64–85.
- Gardner, J. (2012). Assessment and learning: Introduction. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 1–5). London, England: Sage.
- Gattullo, F. (2000). Formative assessment in ELT primary (elementary) classrooms: An Italian case study. *Language Testing*, 17(2), 278–288.
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge, England: Cambridge University Press.
- Gil, J. (2016). English language education policies in the People's Republic of China. In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 49–90). Cham, Switzerland: Springer.
- Gilliland, B. (2015). High school teacher perspectives and practices: Second language writing and language development. *Language and Education*, 29(4), 287–301.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers.

- Profile: Issues in Teachers' Professional Development*, 20(1), 179–195.
- Glasson, T. (2009). *Improving student achievement: A practical guide to assessment for learning*. Carlton, Victoria: Curriculum Corporation.
- Golombek, P. R., & Johnson, K. E. (2004). Narrative inquiry as a mediational space: Examining emotional and cognitive dissonance in second-language teachers' development. *Teachers and Teaching: Theory and Practice*, 10(3), 307–327.
- Goodrich, H. (1996). *Student self-assessment: At the intersection of metacognition and authentic assessment* (Unpublished doctoral dissertation). Harvard University, Cambridge, MA.
- Graden, E. C. (1996). How language teachers' beliefs about reading instruction are mediated by their beliefs about students. *Foreign Language Annals*, 29(3), 387–395.
- Grossman, P. L. (1991). Overcoming the apprenticeship of observation in teacher education coursework. *Teaching and Teacher Education*, 7(4), 345–357.
- 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). Dordrecht, the Netherlands: 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 and Practice*, 21(3), 286–305.
- Guadu, Z. B., & Boersma, E. J. (2018). EFL instructors' beliefs and practices of formative assessment in teaching writing. *Journal of Language Teaching and Research*, 9(1), 42.
- Guba, E. G. (1990). *The paradigm dialog*. Newbury Park, CA: Sage.

- Hallam, S., & Ireson, J. (2003). Secondary school teachers' attitudes towards and beliefs about ability grouping. *British Journal of Educational Psychology*, 73(3), 343–356.
- Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 487–504). Norwell, MA: Springer.
- Harlen, W. (2006a). *Helping learning: A framework for decisions* (4th ed.). London, England: Paul Chapman.
- Harlen, W. (2006b). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp. 103–117). London, England: Sage.
- Harlen, W. (2006c). The role of assessment in developing motivation for learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 61–80). London, England: Sage.
- Harlen, W. (2007). Formative classroom assessment in science and mathematics. In J. H. McMillian (Ed.), *Formative classroom assessment: Theory into practice* (pp. 116–135). New York, NY: Teacher College Press.
- Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy and Practice*, 10(2), 169–207.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy and Practice*, 4(3), 365–379.
- Hasselgreen, A. (2000). The assessment of the English ability of young learners in Norwegian schools: An innovative approach. *Language Testing*, 17(2), 261–277.
- Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing*, 22(3), 337–354.

- Hasselgreen, A. (2012). Assessing young learners. In G. Fulcher & F. Davison (Eds.), *The Routledge handbook of language testing* (pp. 93–105). New York, NY: Routledge.
- Hattie, J., & Jaeger, R. (1998). Assessment and classroom learning: A deductive approach. *International Journal of Phytoremediation*, 21(1), 111–122.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hawe, E., & Parr, J. (2014). Assessment for learning in the writing classroom: An incomplete realisation. *The Curriculum Journal*, 25(2), 210–237.
- Heritage, M. (2013). Using assessment information in real time. In R. W. Lissitz (Ed.), *Informing the practice of teaching using formative and interim assessment: A systems approach* (pp. 85–104). Charlotte, NC: Information Age Publishing.
- Herman, J. L., & Baker, E. L. (2005). Making benchmark testing work. *Educational Leadership*, 63(3), 48–54.
- Hill, K. (2012). *Classroom-based assessment in the school foreign language classroom*. New York, NY: Peter Lang.
- Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420.
- Hipp, K. K., & Huffman, J. B. (2007). Using assessment tool as frames for dialogue to create and sustain professional learning communities. In L. Stoll & K. S. Louis (Eds.), *Professional learning communities: Divergence, depth and dilemmas* (pp. 119–131). New York, NY: Open University Press.
- Hoffman, J., Assaf, L., & Paris, S. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, 54(5), 482–494.

- Hopkins, D. (2014). *A teacher's guide to classroom research* (5th ed.). Maidenhead, UK: Open University Press.
- Howard, T. C. (2003). Culturally relevant pedagogy: Ingredients. *Theory Into Practice*, 42(3), 195–202.
- Hu, G. (2002). Potential cultural resistance to pedagogical imports: The case of communicative language teaching in China. *Language, Culture and Curriculum*, 15(2), 93–105.
- Huberman, M. (1989). The professional life cycle of teachers. *Teacher College Record*, 91, 31–57.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, England: Cambridge University Press.
- Hung, Y. J., Samuelson, B. L., & Chen, S. C. (2016). The relationship between peer- and self-assessment and teacher assessment of young EFL learners' oral presentations. In M. Nikolov (Ed.), *Assessing Young Learners of English: Global and Local Perspectives* (pp. 317–338). New York, NY: Springer.
- Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy and Practice*, 16(2), 185–204.
- Jacobs, G. M., & Renandya, W. A. (2019). *Student centered cooperative learning: Linking concepts in education to promote student learning*. Basingstoke, UK: Springer.
- Jalongo, M. R. (2000). *Early childhood language arts*. Boston, MA: Allyn and Bacon.
- James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 47–60). Thousand Oaks, CA: Sage.
- James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices

- and values. *The Curriculum Journal*, 17(2), 109–138.
- Jin, L., & Cortazzi, M. (1998). Dimensions of dialogue: Large classes in China. *International Journal of Educational Research*, 29(8), 739–761.
- Jin, L., & Cortazzi, M. (2002). English language teaching in China: A bridge to the future. *Asia Pacific Journal of Education*, 22(2), 53–64.
- John-Steiner, V., & Mahn, H. (1996). Sociocultural approaches to learning and development: A Vygotskian framework. *Educational Psychologist*, 31(3/4), 191–206.
- Johnson, K. E. (1992). The relationship between teachers' beliefs and practices during literacy instruction for non-native speakers of English. *Journal of Literacy Research*, 24(1), 83–108.
- Johnson, K. E. (1994). The emerging beliefs and instructional practices of preservice English as a second language teachers. *Teaching and Teacher Education*, 10(4), 439–452.
- Jones, F., & Fong, M. (2007). The impact of teachers' beliefs and educational experiences on EFL classroom practices in secondary schools. *Asian Journal of English Language Teaching*, 17(1), 27–48.
- Junqueira, L., & Kim, Y. (2013). Exploring the relationship between training, beliefs, and teachers' corrective feedback practices: A case study of a novice and an experienced ESL teacher. *Canadian Modern Language Review*, 69(2), 181–206.
- Kagan, D. M. (1990). Ways of evaluating teacher cognition. *Review of Educational Research*, 60(3), 419–469.
- Kagan, D. M. (1992). Implications of research on teacher belief. *Educational Psychologist*, 27(1), 65–90.
- Kahn, E. A. (2000). A case study of assessment in a Grade 10 English course. *The*

- Journal of Educational Research*, 93(5), 276–286.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36.
- Kane, R., Sandretto, S., & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research*, 72(2), 177–228.
- Kang, Y., & Cheng, X. (2014). Teacher learning in the workplace: A study of the relationship between a novice EFL teacher's classroom practices and cognition development. *Language Teaching Research*, 18(2), 169–186.
- Kay, D., & Kibble, J. (2016). Learning theories 101: Application to everyday teaching and scholarship. *Advances in Physiology Education*, 40(1), 17–25.
- Kirkpatrick, A., & Liddicoat, A. J. (2017). Language education policy and practice in East and Southeast Asia. *Language Teaching*, 50(2), 155–188.
- Kiss, C., & Nikolov, M. (2005). Developing, piloting, and validating an instrument to measure young learners' aptitude. *Language Learning*, 55(1), 99–150.
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice*, 16(3), 263–268.
- Kline, R. B. (2011). *Principles and practices of structural equation modeling* (3rd ed.). New York, NY: Guilford.
- Klinger, D. A., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, 76(3), 1–35.
- Kvale, S., & Brinkmann, S. (2015). *Interviews: Learning the craft of qualitative research* (3rd ed.). Los Angeles, CA: Sage.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197.

- Lao, C. (2004). Parents' attitudes toward Chinese–English bilingual education and Chinese-language use. *Bilingual Research Journal*, 28(1), 99–121.
- Lapadat, J. C., & Lindsay, A. C. (1999). Transcription in research and practice: From standardization of technique to interpretive positionings. *Qualitative Inquiry*, 5(1), 64–86.
- Lee, I. (2007a). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *The Canadian Modern Language Review*, 64(1), 199–214.
- Lee, I. (2007b). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning? *Assessing Writing*, 12(3), 180–198.
- Lee, I. (2016). Putting students at the centre of classroom L2 writing assessment. *Canadian Modern Language Review*, 72(2), 258–280.
- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34–50.
- Lee, I., Mak, P., & Burns, A. (2016). EFL teachers' attempts at feedback innovation in the writing classroom. *Language Teaching Research*, 20(2), 248–269.
- Lee, I., Mak, P., & Yuan, R. E. (2019). Assessment as learning in primary writing classrooms: An exploratory study. *Studies in Educational Evaluation*, 62, 72–81.
- Leng, H. (2005). Chinese cultural schema of education: Implications for communication between Chinese students and Australian educators. *Issues in Educational Research*, 15(1), 17–36.
- Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and Change. *Language Assessment Quarterly: An International Journal*, 1(1), 19–41.
- Leung, C. (2005). Classroom teacher assessment of second language development:

- Construct as practice. In E. Hinkel (Ed.), *The handbook of research in second language teaching and learning* (pp. 526–537). Mahwah, NJ: Erlbaum.
- Li, L. (2017). *Social interaction and teacher cognition*. Edinburgh, UK: Edinburgh University Press.
- Li, L. (2020). *Language teacher cognition: A sociocultural perspective*. London, England: Palgrave.
- Liao, X. (2004). The need for communicative language teaching in China. *ELT Journal*, 58(3), 270–273.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Liu, J., & Pan, M. (2019). English language teaching in China: Developing language proficiency frameworks. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 415–432). Cham, Switzerland: Springer.
- Liu, N., Lin, C. K., & Wiley, T. G. (2016). Learner views on English and English language teaching in China. *International Multilingual Research Journal*, 10(2), 137–157.
- Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Lok, B., McNaught, C., & Young, K. (2016). Criterion-referenced and norm-referenced assessments: Compatibility and complementarity. *Assessment and Evaluation in Higher Education*, 41(3), 450–465.
- Looney, J. (2005). *Formative assessment: Improving learning in secondary classrooms*. Paris, France: Organisation for Economic Cooperation and Development.
- López-Pastor, V., & Sicilia-Camacho, A. (2017). Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment and*

- Evaluation in Higher Education*, 42(1), 77–97.
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.
- Lynch, B. K. (1996). *Language program evaluation: Principles and practices*. Cambridge, England: Cambridge University Press.
- MacDonald, M., Badger, R., & White, G. (2001). Changing values: What use are theories of language learning and teaching? *Teaching and Teacher Education*, 17(8), 949–963.
- Mak, P., & Lee, I. (2014). Implementing assessment for learning in L2 writing: An activity theory perspective. *System*, 47, 73–87.
- Mason, J. (2006). Mixing methods in a qualitatively driven way. *Qualitative Research*, 6(1), 9–25.
- Mavrommatis, Y. (1997). Understanding assessment in the classroom: Phases of the assessment process — the assessment episode. *Assessment in Education: Principles, Policy and Practice*, 4(3), 381–400.
- McCusker, K., & Gunaydin, S. (2015). Research using qualitative, quantitative or mixed methods and choice based on the research. *Perfusion*, 30(7), 537–542.
- McDonough, J., & McDonough, S. (1997). *Research methods for English language teachers*. London, England: Routledge.
- McKay, P. (2006). *Assessing young learners*. Cambridge, England: Cambridge University Press.
- McMillan, J. H. (1997). *Classroom assessment: Principles and practice for effective instruction*. Needham Heights, MA: Allyn and Bacon.
- 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.
- McMillan, J. H. (2008). *Assessment essentials for standards-based education* (2nd ed.). Thousand Oaks, CA: Corwin.
- McMillan, J. H. (2010). The practical implications of educational aims and contexts for formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *The handbook of formative assessment* (pp. 41–58). New York, NY: Routledge.
- McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction* (5th ed.). Boston, MA: Pearson.
- McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *The Sage handbook of research on classroom assessment* (pp. 2–17). Thousand Oaks, CA: Sage.
- McMillan, J. H. (2017). *Classroom assessment: Principles and practice that enhance student learning and motivation* (7th ed.). New York, NY: Pearson.
- McNamara, T. (2001a). Language assessment as social practice: Challenges for research. *Language Testing*, 18(4), 333–349.
- McNamara, T. (2001b). Rethinking alternative assessment. *Language Testing*, 18(4), 329–332.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.
- Merriam, S. B. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). San Francisco, CA: Jossey-Bass.
- Meschede, N., Fiebranz, A., Möller, K., & Steffensky, M. (2017). Teachers' professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. *Teaching and Teacher Education*, 66, 158–170.

- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.
- Ministry of Education. (2001a). *English curriculum standards for 9-year compulsory education and senior secondary schools (Trial version)*. Beijing, China: Beijing Normal University Press.
- Ministry of Education. (2001b). *Guidelines for promoting teaching English courses in primary Schools*. China: Document issued by the Ministry of Education.
- Ministry of Education. (2003). *English curriculum standards for senior secondary schools (Trial version)*. Beijing, China: Beijing Normal University Press.
- Ministry of Education. (2011). *English curriculum standards for compulsory education (2011 version)*. Beijing, China: Beijing Normal University Press.
- Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories* (3rd ed.). London, England: Routledge.
- Moodie, I. (2016). The anti-apprenticeship of observation: How negative prior language learning experience influences English language teachers' beliefs and practices. *System*, *60*, 29–41.
- Morgan, D. L. (2014). *Integrating qualitative and quantitative methods: A pragmatic approach*. Thousand Oaks, CA: Sage.
- Mori, R. (2011). Teacher cognition in corrective feedback in Japan. *System*, *39*, 451–467.
- Morris, P. J. T. F., Lo, M. L. F., Chik, P. M. P., & Chan, K. K. (2000). One function, two systems: Changing assessment in Hong Kong's primary schools. In B. Adamson, T. Kwan, & K. K. Chan (Eds.), *Changing the curriculum: The impact of reform on primary schooling in Hong Kong* (pp. 195–215). Hong Kong, China: Hong Kong University Press.

- Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment in every classroom: A guide for instructional leaders*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mui So, W. W., & Hoi Lee, T. T. (2011). Influence of teachers' perceptions of teaching and learning on the implementation of assessment for learning in inquiry study. *Assessment in Education: Principles, Policy and Practice*, 18(4), 417–432.
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile*, 14(1), 143–158.
- Murtagh, L. (2014). The motivational paradox of feedback: Teacher and student perceptions. *The Curriculum Journal*, 25(4), 516–541.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22(2), 155–175.
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19(4), 317–328.
- Nguyen, P., Terlouw, C., & Pilot, A. (2006). Culturally appropriate pedagogy: The case of group learning in a Confucian Heritage Culture context. *Intercultural Education*, 17(1), 1–19.
- Nicol, D., & 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.
- Nishino, T. (2012). Modeling teacher beliefs and practices in context: A multimethods approach. *Modern Language Journal*, 96(3), 380–399.
- Numrich, C. (2006). On becoming a language teacher: Insights from diary studies. *TESOL Quarterly*, 30(1), 131–152.
- Nunan, D., & Bailey, K. M. (2009). *Exploring second language classroom research: A*

- comprehensive guide*. Boston, MA: Heinle/Cengage.
- Önalan, O., & Karagül, A. E. (2018). A study on Turkish EFL teachers' beliefs about assessment and its different uses in teaching English. *Journal of Language and Linguistic Studies*, *14*(3), 190–201.
- Ormrod, J. E. (2014). *Educational psychology: Developing learners* (8th ed.). Boston, MA: Pearson.
- Osborne, J. W., & Banjanovic, E. S. (2016). *Exploratory factor analysis with SAS*. Cary, NC: SAS Institute.
- Öz, H. (2014). Turkish teachers' practices of assessment for learning in the English as a foreign language classroom. *Journal of Language Teaching and Research*, *5*(4), 775–785.
- Packer, M. J., & Goicoechea, J. (2000). Sociocultural and constructivist theories of learning: Ontology, not just epistemology. *Educational Psychologist*, *35*(4), 227–241.
- Pajares, F. M. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*(3), 307–332.
- Pallant, J. (2011). *SPSS Survival Manual: A step by step guide to data analysis using SPSS* (4th ed.). Crows Nest, NSW: Allen and Unwin.
- Pan, M., & Wu, X. (2019). 中国英语能力等级量表在中小学英语形成性评价中的应用——以写作能力为例. *Foreign Language World*, (1), 89–96.
- Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology*, *11*(2), 551–576.
- Panadero, E., Andrade, H., & Brookhart, S. M. (2018). Fusing self-regulated learning

- and formative assessment: A roadmap of where we are, how we got here, and where we are going. *Australian Educational Researcher*, 45(1), 13–31.
- Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self- and co-regulated learning: The role of evaluative judgement. *European Journal of Psychology of Education*, 34(3), 535–557.
- Panadero, E., Jasson, A., & Strijbos, J. W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In L. Allal & D. Laveault (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 311–326). Cham, Switzerland: Springer.
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98.
- Patel, H. R., & Joseph, J. M. (2016). Questionnaire designing process: A review. *Journal of Clinical Trials*, 6(2), 1–7.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Peacock, M. (2001). Pre-service ESL teachers' beliefs about second language learning: A longitudinal study. *System*, 29, 177–195.
- Pedulla, J., Abrams, L., Maduas, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston, MA: National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College.
- Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning

- processes: Towards a wider conceptual field. *International Journal of Phytoremediation*, 21(1), 85–102.
- Phelps, R. P. (2005). The rich, robust research literature on testing's achievement benefits. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 55–90). Mahwah, NJ: Erlbaum.
- Phipps, S., & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System*, 37, 380–390.
- Ping, Q. (2015). Ideologies in primary English textbooks in China. In X. L. Curdt-Christiansen & C. Weninger (Eds.), *Language, ideology and education: The politics of textbooks in language education* (pp. 163–180). London, UK: Routledge.
- Pintrich, P. R., Conley, A. M. M., & Kempler, T. M. (2003). Current issues in achievement goal theory and research. *International Journal of Educational Research*, 39, 319–337.
- Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation*, 38(3), 84–92.
- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, 34(1), 1–20.
- Punch, K. F. (2009). *Introduction to research methods in education*. London, England: Sage.
- Qian, D. (2007). From TOEFL pBT to TOEFL iBT: Recent trends, research landscape, and Chinese learners. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 95–111). New York, NY: Routledge.
- Rahimi, M., & Zhang, L. J. (2015). Exploring non-native English-speaking teachers'

- cognitions about corrective feedback in teaching English oral communication. *System*, 55, 111–122.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13.
- Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics*, 20(1), 18–41.
- Rao, Z. (2013). Teaching English as a foreign language in China: Looking back and forward. *English Today*, 29(3), 34–39.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429–462.
- 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.
- Rea-Dickins, P. (2007). Classroom-based assessment: Possibilities and pitfalls. In C. C. J & Davison (Eds.), *International handbook of English language teaching* (pp. 505–520). New York, NY: Springer.
- Reinholz, D. (2016). The assessment cycle: A model for learning through peer assessment. *Assessment and Evaluation in Higher Education*, 41(2), 301–315.
- Ren, W., & Han, Z. (2016). The representation of pragmatic knowledge in recent ELT textbooks. *ELT Journal*, 70(4), 424–434.
- Ren, W., & Liu, J. (2016). Second language research. *Language Teaching*, 49(2), 295–298.
- Rogers, T., Cheng, L., & Hu, H. (2007). ESL/EFL Instructors' beliefs about assessment and evaluation. *Canadian and International Education/Education*

- Canadienne et Internationale*, 36(1), 39–61.
- Roothoof, H. (2014). The relationship between adult EFL teachers' oral feedback practices and their beliefs. *System*, 46, 65–79.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research and Evaluation*, 11(10), 1–13.
- Rubie-Davies, C. M., Peterson, E., Irving, E., Widdowson, D., & Dixon, R. (2010). Expectations of achievement: Student, teacher and parent perceptions. *Research in Education*, 83(1), 36–53.
- Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37(4), 15–24.
- Russell, M. K., & Airasian, P. W. (2012). *Classroom assessment: Concepts and applications* (6th ed.). Boston, MA: McGraw-Hill.
- Sach, E. (2012). Teachers and testing: An investigation into teachers' perceptions of formative assessment. *Educational Studies*, 38(3), 261–276.
- Sadler, R. D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Saito, H., & Inoi, S. (2017). Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly*, 14(3), 213–233.
- Sato, K., & Kleinsasser, R. C. (2004). Beliefs, practices, and interactions of teachers in a Japanese high school English department. *Teaching and Teacher Education*, 20(8), 797–816.
- Saville, N., & Weir, C. (2018). *Examining young learners: Research and practice in assessing the English of school-age learners*. Cambridge, England: Cambridge

University Press.

- Schrank, F. A., Fletcher, Todd, V., & Alvarado, C. G. (1996). Comparative validity of three English oral language proficiency tests. *Bilingual Research Journal*, 20(1), 55–68.
- Schulte, B. (2017). Private schools in the People’s Republic of China: Development, modalities and contradictions. In T. Koinzer, R. Nikolai, & F. Waldow (Eds.), *Private schools and school choice in compulsory education: Global change and national challenge* (pp. 115–132). Berlin, Germany: Springer.
- Schunk, D. (2000). *Learning theories: An educational perspective*. Upper Saddle River, NJ: Merrill/Prentice-Hall.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives on curriculum evaluation* (pp. 39–83). Chicago: Rand McNally.
- Scrivener, J. (2011). *Learning teaching: The essential guide to English language teaching* (3rd ed.). Oxford: Macmillan.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (pp. 623–646). Westport, CT: Praeger.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21–34.
- Shim, K. N. (2009). *An investigation into teachers’ perceptions of classroom-based assessment of English as a foreign language in Korean primary education*

- (Unpublished doctoral dissertation). The University of Exeter, Korea.
- Shohamy, E. (1994). The role of language tests in the construction of second-language acquisition theories. In E. B. Bernhardt, E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 133–142). Hillsdale, NJ: Lawrence Erlbaum.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4), 331–345.
- Shohamy, E., Inbar-Lourie, O., & Poehner, M. E. (2008). *Investigating assessment perceptions and practices in the advanced foreign language classroom* (Report No. 1108). University Park, PA: Center for Advanced Language Proficiency Education and Research.
- Simons, H. (2009). *Case study research in practice*. London, England: Sage.
- Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford: Oxford University Press.
- Stern, H. H., & Weinrib, A. (1977). Foreign languages for younger children: Trends and assessment. *Language Teaching*, 10(1), 5–25.
- Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Merrill/Prentice-Hall.
- Stiggins, R. J. (2005). From formative assessment to assessment for learning: A path to success in standards. *The Phi Delta Kappan*, 87(4), 324–328.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.
- Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change*,

7(4), 221–258.

- Sun, Q. (2017). *An exploratory study of Chinese university English-as-a-foreign-language (EFL) teachers' cognitions and practices about English grammar teaching (Unpublished doctoral dissertation)*. The University of Auckland, Auckland, NZ.
- Suri, H. (2011). Purposeful sampling in qualitative research synthesis. *Qualitative Research Journal, 11*(2), 63–75.
- Szpotowicz, M., & Campfield, D. E. (2016). Developing and piloting proficiency test for Polish young learners. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 109–137). Cham, Switzerland: Springer.
- Tabachnic, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). New York, NY: Allyn and Bacon.
- Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing. *Language Testing, 17*(2), 163–184.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *The handbook of research on mathematics teaching and learning* (pp. 127–146). New York, NY: Macmillan.
- Timperley, H. S., & Parr, J. M. (2009). What is this lesson about? Instructional processes and student understandings in writing classrooms. *The Curriculum Journal, 20*(1), 43–60.
- Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 55–87). Dordrecht, the Netherlands: Kluwer Academic.

- Topping, K. J. (2013). Peers as a source of formative and summative assessment. In J. H. McMillian (Ed.), *The handbook of research on classroom assessment* (pp. 394–412). Thousand Oaks, CA: Sage.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham: Open University Press.
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615–631.
- Troudi, S., Coombe, C., & Al-hamly, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546–555.
- Tunstall, P., & Gipps, C. (1996a). 'How does your teacher help you to make your work better?' Children's understanding of formative assessment. *The Curriculum Journal*, 7(2), 185–203.
- Tunstall, P., & Gipps, C. (1996b). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, 22(4), 389–404.
- Turner, C. E. (2012). Classroom assessment. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 65–79). London, England: Routledge.
- Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *The handbook of second language assessment* (pp. 1–19). Berlin, Germany: DeGruyter Mouton.
- Urmston, A. (2003). Learning to teach English in Hong Kong: The opinions of teachers in training. *Language and Education*, 17(2), 112–137.
- Vehovar, V., Toepoel, V., & Steinmetz, S. (2016). Non-probability sampling. In C.

- Wolf, D. Joye, T. W. Smith, & Y. Fu (Eds.), *The Sage handbook of survey methodology* (pp. 229–345). Los Angeles, CA: Sage.
- Verloop, N., Van Driel, J., & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research*, *35*, 441–461.
- Vettorel, P. (2014). *English as a lingua franca in wider networking: Blogging practices*. Berlin, Germany: De Gruyter Mouton.
- Vogt, K., & Tzagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374–402.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
- Wang, J. (2020). Exploring the perceived integrations between assessment and metacognition: A qualitative inquiry of three award-winning teacher educators' conceptions of assessment in a Hong Kong university context. *Frontiers in Education*, *4*(157).
- Wang, Q. (2009). Primary English in China: Policy, curriculum and implementation. In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 277–309). Berlin, Germany: Mouton de Gruyter.
- Wang, Q. (2011). On primary English education in China. *Foreign Language of China*, *8*(4), 47–54.
- Wang, X. (2017). A Chinese EFL teacher's classroom assessment practices. *Language Assessment Quarterly*, *14*(4), 312–327.
- Wanner, T., & Palmer, E. (2018). Formative self-and peer assessment for improved student learning: The crucial factors of design, teacher participation and feedback. *Assessment and Evaluation in Higher Education*, *43*(7), 1032–1047.

- Waters, A., & Vilches, M. L. C. (2001). Implementing ELT innovations: A needs analysis framework. *ELT Journal*, 55(2), 133–141.
- Watson, A. (2015). The problem of grammar teaching: A case study of the relationship between a teacher's beliefs and pedagogical practice. *Language and Education*, 29(4), 332–346.
- Wen, Q. (2012). Teaching English as an international language in mainland China. In A. Kirkpatrick & R. Sussex (Eds.), *English as an international language in Asia: Implications for language education* (pp. 79–93). Berlin, Germany: Springer.
- Wiggins, G. (2012). Seven keys to effective feedback. *Educational Leadership*, 70(1), 10–16.
- William, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1053–1098). Greenwich, CT: Information Age.
- William, D. (2010a). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18–40). New York, NY: Routledge.
- William, D. (2010b). The role of formative assessment in effective learning environments. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning using research to inspire practice* (pp. 135–155). Paris, France: OECD Publishing.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14.
- William, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In

- J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 29–42). New York, NY: Teachers College Press.
- William, D., & Thompson, M. (2008). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Erlbaum.
- Woods, D. (1996). *Teacher cognition in language teaching*. Cambridge, England: Cambridge University Press.
- Wu, D., Zhang, L. J., & Wei, L. (2019). Developing translator competence: Understanding trainers' beliefs and training practices. *Interpreter and Translator Trainer, 13*(3), 233–254.
- Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System, 50*, 56–68.
- Xu, Y. (2016). Assessment planning within the context of university English language teaching (ELT) in China: Implications for teacher assessment literacy. *Australian Review of Applied Linguistics, 39*(3), 233–254.
- Xu, Y., & Harfitt, G. (2018). 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.
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly, 43*(3), 492–513.
- Yang, T. L. (2012). Tertiary EFL teachers' classroom assessment practices. *International Journal of Learning, 18*(3), 1–8.
- Yang, W., & Dai, W. (2011). Rote memorization of vocabulary and vocabulary

- development. *English Language Teaching*, 4(4), 61–64.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Los Angeles, CA: Sage.
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Los Angeles, CA: Sage.
- Yoon, E., & Lee, H. K. (2013). Do effects of self-assessment differ by L2 language level? A case of Korean learners of English. *Language Testing*, 22(4), 731–739.
- Young, J. E. J., & Jackman, M. G. A. (2014). Formative assessment in the Grenadian lower secondary school: Teachers' perceptions, attitudes and practices. *Assessment in Education: Principles, Policy and Practice*, 21(4), 398–411.
- Yu, G., & Jin, Y. (2014). English language assessment in China: Policies, practices and impacts. *Assessment in Education: Principles, Policy and Practice*, 21(3), 245–250.
- Yu, G., & Jin, Y. (2016). *Assessing Chinese learners of English: The language constructs, consequences and conundrums*. London, England: Palgrave.
- Yu, S., Lee, I., & Mak, P. (2016). Revisiting Chinese cultural issues in peer feedback in EFL writing: Insights from a multiple case study. *Asia-Pacific Education Researcher*, 25(2), 295–304.
- Yuan, S., & Shu, D. (2017). 我国外语教学中的形成性评价研究：回顾与思考 (2002-2016). *Foreign Language Teaching Theories and Practices*, 4, 51–56.
- Zangl, R. (2000). Monitoring language skills in Austrian primary (elementary) schools: A case study. *Language Testing*, 17(2), 250–260.
- Zeichner, K. M., & Tabachnick, B. R. (1985). The development of teacher perspectives: Social strategies and institutional control in the socialization of beginning teachers. *Journal of Education for Teaching*, 11(1), 1–25.

- 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.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323–342.
- 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.
- Zhou, J., & Deneen, C. C. (2016). Chinese award-winning tutors' perceptions and practices of classroom-based assessment. *Assessment and Evaluation in Higher Education*, 41(8), 1144–1158.
- Zhu, H. (2003). Globalization and new ELT challenges in China. *English Today*, 19(4), 36–41.

APPENDICES

Appendix 1 Primary School English Teachers' Classroom-based Assessment: Conceptions and Practices Questionnaire (English version)

Dear teachers,

This questionnaire is to gain an understanding of Chinese EFL teachers' conceptions and practices regarding classroom assessment of young EFL learners. Please respond to each item by filling the blanks or ticking (√) the appropriate options. Please *choose only one answer for all the multiple-choice questions except for those particularly labeled*. There are no right or wrong answers. *All the information you provide will be treated in confidence and used for research purposes only*. Thanks very much for your support.

Section 1: Your conceptions of classroom assessment purposes

The following statements address the purposes when teachers assess their students in classrooms. Please indicate your opinion by using the following rating scale:

1=Strongly disagree

2=Disagree

3=Somewhat agree

4=Agree

5=Agree very much

6=Completely agree

1	Teachers use assessments to determine how much students have learned from teaching	1	2	3	4	5	6
2	Teachers use assessments to establish what students have learned	1	2	3	4	5	6
3	Teachers use assessments to identify students' strengths and weaknesses	1	2	3	4	5	6
4	Teachers provide feedback to students about their performance	1	2	3	4	5	6
5	Teachers use assessments to identify students' learning needs	1	2	3	4	5	6
6	Teachers use assessments to motivate students to learn	1	2	3	4	5	6
7	Teachers use assessments to help students improve their learning	1	2	3	4	5	6
8	Teachers place students into high proficiency and low proficiency groups according to their assessment performance	1	2	3	4	5	6
9	Teachers assign scores or grades to student work	1	2	3	4	5	6
10	Teachers use assessments to determine if students have met the desired standards	1	2	3	4	5	6

11	Assessment is integrated with teachers' daily teaching practice	1	2	3	4	5	6
12	Teachers diagnose strengths and weaknesses in their teaching through assessment	1	2	3	4	5	6
13	Teachers modify their ongoing teaching based on assessment results						
14	Teachers allow different students to get different instruction based on assessment results	1	2	3	4	5	6
15	Teachers teach according to public examinations' requirements	1	2	3	4	5	6
16	The priority of teachers' work is to help students to pass their examinations	1	2	3	4	5	6
17	Teachers teach students examinations skills	1	2	3	4	5	6
18	Assessment provides information on how well teachers are doing with reference to schools' requirements	1	2	3	4	5	6
19	Assessment results are an indicator of a teacher's teaching quality	1	2	3	4	5	6
20	Assessment results are a way to measure the quality of a school	1	2	3	4	5	6

Section 2 Your conceptions of classroom-based assessment processes

The following statements address the processes that teachers use when assessing their students in classrooms. Please indicate your opinion by using the following rating scale:

1=Not important at all 2=Not important 3=Somewhat important
4=Important 5=Very important 6= Completely important

1	Teachers should identify instructional objectives when designing language activities	1	2	3	4	5	6
2	Teachers establish instructional objectives according to the curriculum requirements	1	2	3	4	5	6
3	Teachers establish instructional objectives according to students' needs	1	2	3	4	5	6
4	Teachers help students understand the objectives	1	2	3	4	5	6
5	Teachers should establish success criteria when planning language activities	1	2	3	4	5	6
6	Teachers help students understand success criteria	1	2	3	4	5	6
7	Teachers select appropriate assessment methods according to students' needs	1	2	3	4	5	6
8	Teachers collect evidence of learning through other teachers	1	2	3	4	5	6
9	Teachers collect evidence of learning through students themselves	1	2	3	4	5	6
10	Teachers collect evidence of learning through students' peers	1	2	3	4	5	6
11	Teachers collect evidence of learning through parents	1	2	3	4	5	6
12	Teachers collect evidence of learning through classroom observation	1	2	3	4	5	6
13	Teachers collect evidence of learning through oral questioning						

14	Teachers collect evidence of learning through conferences with students after class	1	2	3	4	5	6
15	Teachers collect evidence of learning through student portfolios	1	2	3	4	5	6
16	Teachers collect evidence of learning through self-assessment	1	2	3	4	5	6
17	Teachers collect evidence of learning through peer assessment	1	2	3	4	5	6
18	Teachers collect evidence of learning through students' oral presentation (e.g., duty report)	1	2	3	4	5	6
19	Teachers collect evidence of learning through games	1	2	3	4	5	6
20	Teachers collect evidence of learning through role-play activities	1	2	3	4	5	6
21	Teachers collect evidence of learning through classroom tests						
22	Teachers collect evidence of learning through diction	1	2	3	4	5	6
23	Teachers collect evidence of learning through exercises	1	2	3	4	5	6
24	Teachers elicit evidence of learning through reading aloud and reciting	1	2	3	4	5	6
25	Teachers take account of language knowledge (vocabulary, grammar) when interpreting assessment data						
26	Teachers take account of language skills (listening, speaking, reading and writing) when interpreting assessment data	1	2	3	4	5	6
27	Teachers take account of students' learning approaches when interpreting assessment data	1	2	3	4	5	6
28	Teachers take account of students' affective attitudes (interests, attitudes) when interpreting assessment data	1	2	3	4	5	6
29	Teachers compare students' current performance against the pre-set learning objectives when interpreting assessment data.	1	2	3	4	5	6
30	Teachers compare students' current performance against their previous performance when interpreting assessment data.	1	2	3	4	5	6
31	Teachers check the trustworthiness of judgments when interpreting assessment data.	1	2	3	4	5	6
32	Teachers look for overall patterns of students' learning when interpreting assessment data.	1	2	3	4	5	6
33	Teachers provide feedback to students about current achievement through scores and grades	1	2	3	4	5	6
34	Teachers provide feedback to students about current achievement through written comments	1	2	3	4	5	6
35	Teachers provide feedback to students about current achievement through oral comments	1	2	3	4	5	6
36	Teachers communicate with parents about current achievement of students	1	2	3	4	5	6
37	Teachers provide feedback to help students understand their strengths in learning	1	2	3	4	5	6
38	Teachers provide feedback to help students understand their weaknesses in relation to the learning objectives	1	2	3	4	5	6
39	Teachers provide feedback to help students develop error detection strategies	1	2	3	4	5	6

40	Teachers provide feedback to help students find ways of solving problems they have in their learning	1	2	3	4	5	6
41	Teachers provide feedback to help students understand what need to do to improve their work	1	2	3	4	5	6
42	Teachers provide feedback to help students become self-regulated in their learning (e.g., to be aware of their learning objectives)	1	2	3	4	5	6

Section 3 Your classroom assessment practices

How often do you use the following strategies when assessing your students? Please indicate your situation by using the following rating scale:

1=Never 2=Very rarely 3=Rarely 4=Occasionally
5=Frequently 6=Always

1	I identify instructional objectives when designing language activities	1	2	3	4	5	6
2	I establish instructional objectives according to the curriculum requirements	1	2	3	4	5	6
3	I establish instructional objectives according to students' needs	1	2	3	4	5	6
4	I help students understand the objectives	1	2	3	4	5	6
5	I establish success criteria when planning language activities	1	2	3	4	5	6
6	I help students understand success criteria	1	2	3	4	5	6
7	I select appropriate assessment methods according to students' needs	1	2	3	4	5	6
8	I collect evidence of learning through other teachers	1	2	3	4	5	6
9	I collect evidence of learning through students themselves	1	2	3	4	5	6
10	I collect evidence of learning through students' peers	1	2	3	4	5	6
11	I collect evidence of learning through parents	1	2	3	4	5	6
12	I collect evidence of learning through classroom observation	1	2	3	4	5	6
13	I collect evidence of learning through oral questioning						
14	I collect evidence of learning through conferences with students after class	1	2	3	4	5	6
15	I collect evidence of learning through student portfolios	1	2	3	4	5	6
16	I collect evidence of learning through self-assessment	1	2	3	4	5	6
17	I collect evidence of learning through peer assessment	1	2	3	4	5	6
18	I collect evidence of learning through students' oral presentation (e.g., duty report)	1	2	3	4	5	6
19	I collect evidence of learning through games	1	2	3	4	5	6
20	I collect evidence of learning through role-play activities	1	2	3	4	5	6
21	I collect evidence of learning through classroom tests						
22	I collect evidence of learning through diction	1	2	3	4	5	6
23	I collect evidence of learning through exercises	1	2	3	4	5	6

24	I elicit evidence of learning through reading aloud and reciting	1	2	3	4	5	6	
25	I take account of language knowledge (vocabulary, grammar) when interpreting assessment data							
26	I take account of language skills (listening, speaking, reading and writing) when interpreting assessment data	1	2	3	4	5	6	
27	I take account of students' learning approaches when interpreting assessment data	1	2	3	4	5	6	
28	I take account of students' affective attitudes (interests, attitudes) when interpreting assessment data	1	2	3	4	5	6	
29	I compare students' current performance against the pre-set learning objectives when interpreting assessment data.	1	2	3	4	5	6	
30	I compare students' current performance against their previous performance when interpreting assessment data.	1	2	3	4	5	6	
31	I check the trustworthiness of judgments when interpreting assessment data.	1	2	3	4	5	6	
32	I look for overall patterns of students' learning when interpreting assessment data.		1	2	3	4	5	6
33	I provide feedback to students about current achievement through scores and grades		1	2	3	4	5	6
34	I provide feedback to students about current achievement through written comments		1	2	3	4	5	6
35	Teachers provide feedback to students about current achievement through oral comments		1	2	3	4	5	6
36	I communicate with parents about current achievement of students		1	2	3	4	5	6
37	I provide feedback to help students understand their strengths in learning		1	2	3	4	5	6
38	I provide feedback to help students understand their weaknesses in relation to the learning objectives		1	2	3	4	5	6
39	I provide feedback to help students develop error detection strategies		1	2	3	4	5	6
40	I provide feedback to help students find ways of solving problems they have in their learning		1	2	3	4	5	6
41	I provide feedback to help students understand what need to do to improve their work		1	2	3	4	5	6
42	I provide feedback to help students become self-regulated in their learning (e.g., to be aware of their learning objectives)		1	2	3	4	5	6

Section 4 Demographic information

1. Gender: Male Female

2. Age:

- under 20 years old 21-30 years old 31-40 years old 41-50 years old
over 50 years old

3. Educational qualification: Below Bachelor Bachelor Master Doctoral

4. You have been teaching English at primary school level for:
within 1 year 1-5 years 6-10 years 11-15 years 16-20 years more than 20 years
5. You are now teaching in a: stated funded primary school private primary school
6. You are now teaching (*you may choose more than one*):
Grade 1 Grade 2 Grade 3 Grade 4 Grade 5 Grade 6
7. Number of English classes you are teaching: _____
8. Total number of English lessons you are teaching per week: _____
9. Have you received any course or training in assessment? Yes No

This is the end of the questionnaire. If you are willing to participate in our follow-up classroom observation and interview, please provide your contact information (Email): _____

Appendix 2 Primary School English Teachers' Classroom Assessment: Conceptions and Practices Questionnaire (Chinese version)

小学英语教师对课堂评价的认识和实践

尊敬的老师:

您好! 本研究的目的是了解小学英语教师对课堂评价的认识与实践情况。请您填写或选择答案(√), 除特别标注外, 选择题均为单选。您的回答没有对错之分, 本问卷调查结果只做研究用, 我们将对您所提供的信息保密。非常感谢您的支持与帮助!

第一部分 您对课堂评价目的的认识

下列关于教师在日常教学中评价学生的目的的说法, 您是如何认为的?

	1=完全不同意	2=很不同意	3=不大同意	4=有点同意	5=很同意	6=极其同意
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						

20 对学生的评价结果是考核学校办学水平的一种途径 1 2 3 4 5 6

第二部分 您对课堂评价过程的认识

下列关于教师在日常教学中评价学生的做法，您是如何认为的？

	1=完全不重要 极其重要	2=很不重要	3=不大重要	4=有点重要	5=很重要	6=
1 教师在设计教学活动时明确教学目标是什么	1	2	3	4	5	6
2 教师根据教学大纲要求确定教学目标	1	2	3	4	5	6
3 教师根据学生的需求确定教学目标	1	2	3	4	5	6
4 教师让学生了解教学目标	1	2	3	4	5	6
5 教师在设计教学活动时明确评价标准是什么	1	2	3	4	5	6
6 教师让学生了解评价标准	1	2	3	4	5	6
7 教师在设计教学活动时根据学生情况选择恰当的评价方法	1	2	3	4	5	6
8 教师通过其他任课老师了解学生的学习	1	2	3	4	5	6
9 教师通过学生本人了解学生的学习	1	2	3	4	5	6
10 教师通过学生的同伴了解学生的学习	1	2	3	4	5	6
11 教师通过学生的家长了解学生的学习	1	2	3	4	5	6
12 教师通过课堂上对学生的观察了解学生的学习	1	2	3	4	5	6
13 教师通过课堂提问了解学生的学习	1	2	3	4	5	6
14 教师通过课后与学生谈话了解学生的学习	1	2	3	4	5	6
15 教师通过建立学生学习档案了解学生的学习	1	2	3	4	5	6
16 教师通过学生自评了解学生的学习	1	2	3	4	5	6
17 教师通过同伴评价了解学生的学习	1	2	3	4	5	6
18 教师通过学生个人陈述（如值日报告）了解学生的学习	1	2	3	4	5	6
19 教师通过游戏活动了解学生的学习	1	2	3	4	5	6
20 教师通过角色扮演活动了解学生的学习	1	2	3	4	5	6
21 教师通过课堂测验检查学生的学习	1	2	3	4	5	6
22 教师通过听写检查学生的学习	1	2	3	4	5	6
23 教师通过课后练习题检查学生的学习	1	2	3	4	5	6
24 教师通过朗读和背诵活动检查学生的学习	1	2	3	4	5	6
25 教师在评价学生时关注学生的语言知识（如语法，词汇）	1	2	3	4	5	6
26 教师在评价学生时关注学生的语言技能（听，说，读，写）	1	2	3	4	5	6
27 教师在评价学生时关注学生的学习方法	1	2	3	4	5	6
28 教师在评价学生时关注学生的情感态度（如兴趣，学习态度）	1	2	3	4	5	6
29 教师在评价学生时将学生的学习现状与既定的学习目标作对比	1	2	3	4	5	6

30	教师在评价学生时将学生的学习现状与他之前的情况作对比	1	2	3	4	5	6
31	教师在评价学生时确保评价的真实可信	1	2	3	4	5	6
32	教师在评价学生时总结学生学习情况的特征	1	2	3	4	5	6
33	教师通过分数和等级让学生了解自己取得的成绩	1	2	3	4	5	6
34	教师通过书面评语让学生了解自己取得的成绩	1	2	3	4	5	6
35	教师通过口头评语让学生了解自己取得的成绩	1	2	3	4	5	6
36	教师与家长交流学生取得的成绩	1	2	3	4	5	6
37	教师通过反馈让学生认识到自己学习上的优势	1	2	3	4	5	6
38	教师通过反馈让学生认识到自己与学习目标的差距	1	2	3	4	5	6
39	教师通过反馈让学生学习从错误中总结经验	1	2	3	4	5	6
40	教师通过反馈引导学生解决自己学习中的困难	1	2	3	4	5	6
41	教师通过反馈让学生知道如何采取措施提高自己的学习	1	2	3	4	5	6
42	教师通过反馈让学生学会管理自己的学习（如明确学习目标）	1	2	3	4	5	6

第三部分 您在日常教学中评价学生的实践情况

您在日常教学中评价学生时，进行以下活动的频率？

1=从不 2=极少 3=很少 4=有时 5=经常 6=总是

1	我在设计教学活动时明确教学目标是什么	1	2	3	4	5	6
2	我根据教学大纲要求确定教学目标	1	2	3	4	5	6
3	我根据学生的需求确定教学目标	1	2	3	4	5	6
4	我让学生了解教学目标	1	2	3	4	5	6
5	我在设计教学活动时明确评价标准是什么	1	2	3	4	5	6
6	我让学生了解评价标准	1	2	3	4	5	6
7	我在设计教学活动时根据学生情况选择恰当的评价方法	1	2	3	4	5	6
8	我通过其他任课老师了解学生的学习	1	2	3	4	5	6
9	我通过学生本人了解学生的学习	1	2	3	4	5	6
10	我通过学生的同伴了解学生的学习	1	2	3	4	5	6
11	我通过学生的家长了解学生的学习	1	2	3	4	5	6
12	我通过课堂上对学生的观察了解学生的学习	1	2	3	4	5	6
13	我通过课堂提问了解学生的学习	1	2	3	4	5	6
14	我通过课后与学生谈话了解学生的学习	1	2	3	4	5	6
15	我通过建立学生学习档案了解学生的学习	1	2	3	4	5	6
16	我通过学生自评了解学生的学习	1	2	3	4	5	6
17	我通过同伴评价了解学生的学习	1	2	3	4	5	6
18	我通过学生个人陈述（如值日报告）了解学生的学习	1	2	3	4	5	6

19	我通过游戏活动了解学生的学习	1	2	3	4	5	6
20	我通过角色扮演活动了解学生的学习	1	2	3	4	5	6
21	我通过课堂测验检查学生的学习	1	2	3	4	5	6
22	我通过听写检查学生的学习	1	2	3	4	5	6
23	我通过课后练习题检查学生的学习	1	2	3	4	5	6
24	我通过朗读和背诵活动检查学生的学习	1	2	3	4	5	6
25	我在评价学生时关注学生的语言知识（如语法，词汇）	1	2	3	4	5	6
26	我在评价学生时关注学生的语言技能（听，说，读，写）	1	2	3	4	5	6
27	我在评价学生时关注学生的学习方法	1	2	3	4	5	6
28	我在评价学生时关注学生的情感态度（如兴趣，学习态度）	1	2	3	4	5	6
29	我在评价学生时将学生的学习现状与既定的学习目标作对比	1	2	3	4	5	6
30	我在评价学生时将学生的学习现状与他之前的情况作对比	1	2	3	4	5	6
31	我在评价学生时确保评价的真实可信	1	2	3	4	5	6
32	我在评价学生时总结学生学习情况的特征	1	2	3	4	5	6
33	我通过分数和等级让学生了解自己取得的成绩	1	2	3	4	5	6
34	我通过书面评语让学生了解自己取得的成绩	1	2	3	4	5	6
35	我通过口头评语让学生了解自己取得的成绩	1	2	3	4	5	6
36	我与家长交流学生取得的成绩	1	2	3	4	5	6
37	我通过反馈让学生认识到自己学习上的优势	1	2	3	4	5	6
38	我通过反馈让学生认识到自己与学习目标的差距	1	2	3	4	5	6
39	我通过反馈让学生学习从错误中总结经验	1	2	3	4	5	6
40	我通过反馈引导学生解决自己学习中的困难	1	2	3	4	5	6
41	我通过反馈让学生知道如何采取措施提高自己的学习	1	2	3	4	5	6
42	我通过反馈让学生学会管理自己的学习（如明确学习目标）	1	2	3	4	5	6

第四部分 您的基本信息

1. 性别：男 女
2. 年龄：20岁及以下 21—30岁 31—40岁 41—50岁 50岁以上
3. 教育背景：本科以下 本科 硕士 博士
4. 教小学英语的时间：1年以内 1-5年 6-10年 11-15年 16-20年 20年以上
5. 所在学校是：公立小学 私立小学
6. 现在所教年级（可多选）：一年级 二年级 三年级 四年级 五年级 六年级
7. 共教_____个英语教学班
8. 每周英语总课时：_____节课

9. 是否参加过关于评价、考试方面的课程或培训? 是 否

本问卷到此结束, 如果您愿意参与后期的访谈和课堂观察, 请留下您的联系方式:

邮箱_____

再次感谢您的参与!

Appendix 3 Interview Guide

Participant _____ Interviewer _____ Date _____

Time _____ School _____

1 Background information

- Gender
- Age
- Educational qualification
- Major
- How long have you been teaching English to primary school students?
- Which grade level are you teaching now?
- Number of classes you are teaching
- Number of lessons you are teaching per week
- Have you ever received any training or attained any course in assessment? If yes, could you provide more information about the content?

2 Questions related to classroom assessment practices

About assessment purposes

- Can you tell me about the purposes that you have for assessing your students? (Will you focus on identifying the strengths and weaknesses of students' learning? Do you use assessments to motivate students to learn? Do you use assessment to know the progress of students' learning?)
- Do you use assessment to check students' attainment? Why do you think it is important to check the attainment?
- Do you use assessment to facilitate your teaching?
- Do you help parents know students' learning performance?

About planning assessment

- Do you plan assessment activities prior to your instruction? Why? (Is assessment a part of your daily teaching? Why?)

- Do you set clear instructional objectives? Why? How do you set these objectives?
- Do you share these objectives with your students? Why? How do you share the objectives?
- Do you set success criteria? Why?
- Do you allow students to know the criteria? Why? How do you share the criteria with students?
- How do you select appropriate assessment methods? Why?
- What are the criteria when you are selecting the assessment methods?

About collecting assessment information

- How do you obtain an understanding of your students' learning? (How about parents, students themselves, peers, other teachers?)
- What methods do you use when you collect evidence about your students' learning? (Do you use classroom tests, exercises, dictation, oral reading? Why?)
- Do you observe your students' performance during the class? Why?
- Do you get to know how your students are doing through oral questions? Why?
- Do you use self-assessment? Why?
- Do you engage students in peer assessment? Why?
- Do you get to know your student' learning through oral presentation activities (role play)? Why?
- Do you get to know your students' learning through games? Why?

About making judgments

- When you make a judgment of students' performance, how do you know they have done a good job or not? (Do you have any criteria? Why?)
- Could you tell me your focus on assessing children' learning in your class? (Did you value children' efforts? Why?)
- Will you check the reliability of the judgment? Why?

About providing feedback

- What type of feedback do you usually provide? (Scores, written comments, oral comments?) Why do you provide this type of feedback?
- Whom do you usually provide the feedback to? (Students themselves, their parents) Why?
- How do you use the evidence of children' learning to plan your lesson? Could you give some examples and explain why you do it in this way?
- How do you provide feedback to promote students' learning? Could you give some examples and explain why you do it in this way?

Other relevant questions

- What kind of difficulties do you face when conducting assessment activities? (Heavy workload, class size, students' attitudes, parents' attitudes, exams, school policy?)
- Is there anything else you want to tell me about your practices of assessment during the class that I haven't asked you?

Appendix 4 Example of Coding Observation Data (Amy's class)

<p>Physical setting: Teacher: Amy Time: 01:40-02:20 28th March, 2017 Unit: Unit3 At the Zoo What's it? Class size:47 The teacher wrote "Unit 3 At the Zoo What's it" on the blackboard. And put on "A,B.C.D" stickers on the blackboard,</p>		
Instructional activity (excerpts of observation data)		Coding
1	<p>Assessment: three minutes oral presentation T: Now, whose turn to 3 minutes before class? S1: (a girl) S: 听不到 T: loudly, loudly, loudly 什么意思? 声音大一点, ok, one two begin. S1: 大家好, 今天我要唱 ABC, ABCDEFG,HIJKLMN,OPQRSTUVWXYZ,XYZ, XYZ, now you see, I can say my ABC T: OK, now, who want to judge her song? Ok, please S2:声音太小了, T: the voice is so small, yes or no? S2:而且他唱的是我们三年级上册学的歌, T: ok, anyone else? S3: 她一直看着你 T: ok, 没有看着你们对不对? T: 3 minutes,下次我们就这样哈, 如果说有同学没有准备, 我们就请两个同学来, 如果你实在没有准备的, you can read the book, or words, or sentences, 你可以读我们学的单词和句子, 当然更希望你抓紧背诵它, 我们页要求了背诵。新增加一个要求, 每</p>	<p>1) 3 minutes oral production: check oral ability 2) Peer judgment (Students were asked to judge the student's performance. But they only provide negative feedback on this student)</p>

	<p>个来参加 3 minutes 的同学，都要讲授一个 new word to others, 每一个 everyone, are you clear? 听清楚了吗， a new word,不是学过的.</p> <p>S:听清楚了。</p>	
2	<p>Assessment: textbook exercises</p> <p>T: now, open your book, 打开你的书。 turn to page 29,</p> <p>T:29, the homework you do, 做了没有?</p> <p>S: 做了。</p> <p>T:你们做的什么?</p> <p>T:没有做的站起来看一下。 给你们一分钟完成 完成了就 sit down, if you finish you can sit down, please</p> <p>(students were doing their homework in class)</p> <p>T: ok, group D is very good, 第一大组没有一个同学没有完成作业的。</p> <p>Check answers</p> <p>T:好了，我们来看一下， look at this, the first one start to read. Look at this one, what's this? It's a?</p> <p>S: fish</p> <p>T:what's this</p> <p>S:mouse</p> <p>T:what's this?</p> <p>S:dog</p> <p>T:what's this?</p> <p>S: panda</p> <p>T:ok, look at this, what color is it?</p> <p>S:red</p> <p>T:it has long?</p>	<p>1) Providing evaluative feedback: Punishment (Students who did not do their homework were asked to stand up)</p> <p>2) Making a judgment: Compare students' work against others' (Students were asked to stand up and all other students could see these students had not finished their work)</p> <p>3) Textbook exercises</p> <p>4) General praise ("Group D is very good")</p>

<p>S:long tail</p> <p>T: it has long tail, very good. Look at this, it has long?</p> <p>S:body</p> <p>T:long body. It has ?</p> <p>S:fat body</p> <p>T:fat body. It has small?</p> <p>S:small eyes</p> <p>T:small eyes?</p> <p>S:small ears.</p> <p>T:small ears. Ok, now, let's listen (the teacher read the sentence by herself). The first one. It's small and thin.</p> <p>Small what's meaning?</p> <p>S:小的,</p> <p>T:thin what's meaning?</p> <p>S:瘦的</p> <p>T:which one is small and thin?</p> <p>S:mouse</p> <p>T:OK, maybe. It's big and fat, big what's meaning?</p> <p>S:大的</p> <p>T: fat, what's meaning?</p> <p>S:肥的</p> <p>T:肥的, 胖的, which one is big and fat</p> <p>S:panda</p> <p>T:maybe panda. Next one, it has long tail and big eyes</p> <p>S: fish</p> <p>T:fish? Fish has long tail? maybe fish. Next one, it has a long body and short legs</p> <p>S:dog</p> <p>T:ok, let's see which one is correct answer. The first one, Ok,正确的同学举手</p>	<p>5) Make a judgment against each other's performance</p>
---	--

	<p>T: are your right?正确没有? If you right, please hands up.</p> <p>T:group A, group B. 每一组都还是完成了。 The next one. It's big and fat. Ok, please hands up.</p> <p>T:ok, some students not.</p> <p>T: the next one. The fish has big eyes and long tail. If you only see long tail, which one or two</p> <p>S:mouse and fish</p> <p>T: mouse and fish good.</p> <p>T:next one, if you right hands up.</p> <p>T: B 大组错的人多,下去之后自己检查</p>	<p>(When the teacher was checking answers to exercises on the book, she asked those who had the correct answers to raise their hands. She even pointed out that students from one group had too many mistakes.)</p> <p>6) Providing evaluative feedback: Disapproval: expression of disappointment (The teacher said one group made too many mistakes)</p>
3	<p>Assessment: oral production</p> <p>T: ask some students to describe some animals and other students to guess, 请一些同学来描述动物, 其他同学来猜, are you clear? You must use colour 你可以使用颜色, or use long, big, fat, short, small, are you clear? Who want to try? Ok, come here.</p> <p>S:老师可以做动作吗</p> <p>T:you can</p> <p>S1: It's small and green</p>	<p>1) Sharing assessment criteria through explanations of the rules</p> <p>2) Providing descriptive feedback: Identify</p>

	<p>T: small and green, any more else? ok, who want to guess?</p> <p>S2: snake</p> <p>S1: no</p> <p>S3: frog</p> <p>S1: yes,青蛙</p> <p>T:ok,这个描述, you can say, it has big eyes. Ok, one point for your group.</p> <p>T: ok, please.</p> <p>S2:老师我可以做动作吗?</p> <p>T: of course.</p> <p>S2:au, au.</p> <p>T:tiger</p> <p>T:You should describe it, 你还是要增加描述的动作。</p> <p>S2: it's yellow, and black. Zhang Yulin</p> <p>S3: Tiger</p> <p>S2:yes.</p> <p>T: you do not use these words, 你没有用到我们才学到的那些词. Ok, others, who want try?</p> <p>T: ok, that girl</p> <p>S3:it has long tail. It's brown and white.</p> <p>T:any more else?还有吗, ok who want to guess?</p> <p>S4:monkey</p> <p>S3:yes</p> <p>T:ok, one point for your group.</p>	<p>weaknesses in students' performance (you can say it has big eyes; you did not use the words we have learned)</p>
4	<p>Assessment: textbook exercises (listen and tick p30)</p>	<p>Textbook exercises</p>

<p>T:ok, turn to page 30. There are two pictures? What's this?</p> <p>S:dog</p> <p>T:this?</p> <p>S:bear</p> <p>T:this?</p> <p>S:panda</p> <p>T:this?</p> <p>S:rabbit</p> <p>T:ok, let's listen. You write which one is true.</p> <p>(played the radio twice: One, it's big and tall, it has small ears, it's brown, what is it?; Two, it has long ears and a short tail, it's white, what is it)</p> <p>T:ok, this sentence, what is it? What's meaning?这句话</p> <p>S: silence</p> <p>T: ok, the first picture, you choose which one? 第一幅图片, 你选择了哪个? dog or bear?</p> <p>S:dog,</p> <p>S:bear, bear</p> <p>T:ok, let's listen again. 我们来听一下。</p> <p>(played the radio again)</p> <p>T:it's big and tall, tall what's meaning?</p> <p>S:高</p> <p>T:it has small ears, small ears</p> <p>S:小小的耳朵</p> <p>T:大的耳朵还是小的?</p> <p>S:小的</p> <p>T:ok, which one is true.</p> <p>S:no.2</p>	<p>Oral question: checking understanding</p>
---	--

	<p>T: Picture two, which one is right?</p> <p>S:rabbit</p> <p>T:它是怎么描述的? describe the rabbit</p> <p>S:it has long ears and it's white</p>	
5	<p>Assessment: Look and match exercises 看图连线</p> <p>T:第二题, look and match.没有做的, 一分钟完成它, 看图连线。 (Students were doing the exercises in class.)</p> <p>T:ok, are you finished it? 完成没? 最后三秒</p> <p>T: ok, let' see. Fat, do you find it?</p> <p>S:兔子</p> <p>T: which animal? Mouse or cat?</p> <p>S:Silence</p> <p>T:Fat, can you find the picture</p> <p>S:cat.</p> <p>T:cat, ok. 它选的什么? 除了 cat 外, 它选了 mouse, 但是我们也可以指哪个地方</p> <p>S:body</p> <p>T:body, ok. Next one, big 它选的是哪个? The cat, we can also use the big,这个猫是大的, 答案不唯一。</p> <p>T: long, what's meaning?</p> <p>S:长的</p> <p>T:长的什么?</p> <p>S:尾巴</p> <p>T:OK, small. 这几个答案都不唯一, 你可以根据你的实际情况来选择。</p>	Textbook exercises: look and match exercises

<p>6</p>	<p>Instruction and assessment: A song</p> <p>T: ok, let's sing. (played the radio: at the zoo, cat is fast. Rat is fat. See them run, fun, fun,fun. Monkey is short. Giraffe is tall. See them play with the ball)</p> <p>T:第二遍时候跟着唱 (play the radio: at the zoo, cat is fast. Rat is fat. See them run, fun, fun,fun. Monkey is short. Giraffe is tall. See them play with the ball)</p> <p>T: ok, 我们来解释下里面的内容。At the zoo, 在? S: 动物园 (cat is fast. Rat is fat) T:ok, cat is fast. Fat what's meaning? S:快的 T: 快的, 请把这个单词的意思写在那里。Please write the meaning on this word. 请把 fast 的意思写在那里, 快的, 迅速的。Ok, rat is fat. 什么是胖的? S: 老鼠。 T: 老鼠, ok (See them run. Fun fun fun) T:See them run, run what's meaning S:跑 T:fun,fun, fun S: 有趣的 T:有趣的, 开心的。The cat and the fat is very happy, yes or no? S:yes (Monkey is short. Giraffe is tall)</p>	<p>1) Oral question: to check whether students were following the instruction</p> <p>2) Observation: (The teacher noticed that only groups C and D were singing the song)</p>
----------	--	---

	<p>T:monkey is short?</p> <p>S:很矮</p> <p>T:猴子是矮的, giraffe is tall.</p> <p>(See them play with the ball)</p> <p>T:see them play with the ball, ball what's meaning?</p> <p>S:球</p> <p>T:now,现在我们听第二遍, 跟着一起唱</p> <p>(played the radio: at the zoo, cat is fast. Rat is fat. See them run, fun, fun,fun. Monkey is short. Giraffe is tall. See them play with the ball)</p> <p>T: ok, once again.我只听到了 DC 组的声音, AB 组声音再大一点</p> <p>(students sing with the radio: at the zoo, cat is fast. Rat is fat. See them run, fun, fun,fun. Monkey is short. Giraffe is tall. See them play with the ball)</p> <p>T:please take out of your hands 拿出你的手来打节拍</p> <p>(students sing with the radio: at the zoo, cat is fast. Rat is fat. See them run, fun, fun,fun. Monkey is short. Giraffe is tall. See them play with the ball)</p> <p>T:ok, 下课之后, 回去有资源的利用起来学一下。这首歌还是比较简单。</p> <p>S: 老师, 有点读机</p>	
7	<p><i>Instruction and assessment</i></p> <p>T:ok, see the story time.</p> <p>(Play the radio)</p> <p>T:ok, 我们再来一句句讲一遍。Zim to find Zoom.</p> <p>Look how many animals do you know? how many animals</p>	<p>Oral question: inform instruction</p>

<p>S:有几个动物</p> <p>T:有几个动物, ok, do you know 你知道几个动物。</p> <p>The next one. I know all the animals. Ok, please</p> <p>S:全部动物</p> <p>T: 全部, OK very good, all 什么意思?</p> <p>S:全部</p> <p>T:Qi , 下课跟着我走。 A tiger and a lion, tiger?</p> <p>S:老虎</p> <p>T:a lion?</p> <p>S:狮子</p> <p>T: a fox and a wolf</p> <p>S:狐狸和狼</p> <p>T: a fox and a wolf, ok. 狐狸和狼。 A zebra and a horse</p> <p>S:斑马和马</p> <p>T: a zebra?</p> <p>S:斑马</p> <p>T:horse</p> <p>S:马</p> <p>T: OK, 这一句 I know they are monkeys. 他们都是猴子, yes or no?</p> <p>S: no.</p> <p>T: the small is a?</p> <p>S: a monkey</p> <p>T:the big one is a?</p> <p>S:silence</p> <p>T:怎么读的这个单词? letter o, 发的是/ou/的音, gorilla,这里什么意思, 猩猩, 大猩猩。 Oops 什么意思?</p> <p>S:糟糕</p>	<p>Punishment</p>
--	-------------------

	<p>T:现在 four students in a group 四人一小组，练习，自己扮演， who want to act Zim, who want to act Zoom,</p> <p>你知道几种动物？ The teacher was walking around. 自己去排练。</p> <p>(Students were preparing for the role play.)</p> <p>T:ok,今天下去后，这里面有些 new words in this storytime.在这个故事环境里面有些新单词，今天下去的任务去了解， read 读这些单词。第二个任务， turn to page 39, 39, 我们的什么任务，下去预习 39 页上面的内容。下一节课。下去后，预习 39 页。下一节课我们将学习这一部分。 Ok, now, 剩下的时间可以预习这部分内容</p>	<p>Assignment: oral reading</p>

Appendix 5 Descriptive Statistics for Items of Purpose Scale

Items	Mini	Max	M	SD	Skewness		Kurtosis	
					Stat	SE	Stat	SE
Q1.1	2	6	4.82	.852	-.658	.174	.263	.346
Q1.2	3	6	4.77	.885	-.486	.174	-.377	.346
Q1.3	2	6	4.86	.900	-.750	.174	.550	.346
Q1.4	3	6	5.17	.634	-.283	.174	-.051	.346
Q1.5	3	6	4.87	.867	-.612	.174	-.116	.346
Q1.6	3	6	5.14	.689	-.577	.174	.579	.346
Q1.7	3	6	5.19	.618	-.275	.174	.105	.346
Q1.8	1	6	3.61	1.344	-.252	.174	-.614	.346
Q1.9	1	6	3.15	1.199	.000	.174	-.311	.346
Q1.10	1	6	4.10	1.072	-.449	.174	.191	.346
Q1.11	3	6	5.02	.749	-.619	.174	.458	.346
Q1.12	3	6	5.08	.762	-.633	.174	.280	.346
Q1.13	3	6	5.16	.767	-.834	.174	.697	.346
Q1.14	3	6	5.23	.681	-.514	.174	.015	.346
Q1.15	1	6	3.94	1.094	-.388	.174	.075	.346
Q1.16	1	6	3.52	1.146	-.130	.174	-.227	.346
Q1.17	2	6	4.24	.842	-.168	.174	.236	.346
Q1.18	2	6	4.50	.922	-.551	.174	.413	.346
Q1.19	1	6	4.09	1.152	-.560	.174	.443	.346
Q1.20	1	6	4.11	1.086	-.460	.174	.256	.346

Appendix 6 Descriptive Statistics for Items of Process Scale

Items	Mini	Max	M	SD	Skewness		Kurtosis	
					Stat	SE	Stat	SE
Q2.1	4	6	5.37	.607	-.395	.174	-.656	.346
Q2.2	3	6	5.25	.691	-.662	.174	.384	.346
Q2.3	3	6	5.11	.713	-.590	.174	.480	.346
Q2.4	3	6	4.84	.753	-.533	.174	.295	.346
Q2.5	3	6	5.07	.704	-.452	.174	.212	.346
Q2.6	2	6	4.76	.884	-.366	.174	-.310	.346
Q2.7	3	6	5.14	.681	-.385	.174	-.077	.346
Q2.8	2	6	4.65	.833	-.068	.174	-.305	.346
Q2.9	3	6	4.96	.710	-.291	.174	-.074	.346
Q2.10	3	6	4.59	.743	.127	.174	-.380	.346
Q2.11	3	6	4.81	.689	.171	.174	-.706	.346
Q2.12	3	6	5.12	.662	-.355	.174	.106	.346
Q2.13	3	6	5.05	.664	-.377	.174	.375	.346
Q2.14	3	6	4.91	.690	-.258	.174	.033	.346
Q2.15	2	6	4.72	.830	-.417	.174	.013	.346
Q2.16	3	6	4.67	.700	-.085	.174	-.186	.346
Q2.17	3	6	4.58	.778	-.168	.174	-.329	.346
Q2.18	2	6	4.59	.714	-.167	.174	.328	.346
Q2.19	3	6	4.70	.743	.097	.174	-.506	.346
Q2.20	3	6	4.73	.718	-.059	.174	-.297	.346
Q2.21	3	6	4.92	.642	-.167	.174	.077	.346
Q2.22	3	6	4.75	.719	-.268	.174	-.016	.346
Q2.23	3	6	4.91	.651	-.133	.174	-.061	.346
Q2.24	3	6	4.89	.653	-.113	.174	-.106	.346
Q2.25	3	6	4.84	.711	-.195	.174	-.146	.346
Q2.26	4	6	5.17	.632	-.150	.174	-.561	.346
Q2.27	3	6	5.11	.661	-.444	.174	.511	.346
Q2.28	3	6	5.20	.647	-.332	.174	-.163	.346
Q2.29	3	6	4.94	.589	-.294	.174	.829	.346

Q2.30	3	6	4.96	.641	-.558	.174	1.195	.346
Q2.31	3	6	5.06	.636	-.290	.174	.364	.346
Q2.32	3	6	4.98	.642	-.455	.174	.897	.346
Q2.33	2	6	4.48	.808	-.326	.174	-.197	.346
Q2.34	3	6	4.68	.741	-.320	.174	-.061	.346
Q2.35	3	6	4.68	.675	-.229	.174	.015	.346
Q2.36	3	6	4.73	.704	-.200	.174	-.065	.346
Q2.37	3	6	5.06	.631	-.297	.174	.431	.346
Q2.38	3	6	5.01	.622	-.133	.174	.052	.346
Q2.39	3	6	5.09	.635	-.200	.174	-.060	.346
Q2.40	3	6	5.08	.645	-.310	.174	.272	.346
Q2.41	3	6	5.10	.642	-.331	.174	.329	.346
Q2.42	3	6	5.09	.640	-.316	.174	.336	.346

Appendix 7 Descriptive Statistics for Items of Practice Scale

Items	Mini	Max	<i>M</i>	<i>SD</i>	Skewness		Kurtosis	
					Stat	SE	Stat	SE
Q3.1	4	6	5.46	.620	-.707	.174	-.464	.346
Q3.2	4	6	5.44	.634	-.694	.174	-.502	.346
Q3.3	3	6	5.25	.726	-.577	.174	-.329	.346
Q3.4	2	6	4.96	.840	-.618	.174	.483	.346
Q3.5	3	6	5.29	.673	-.518	.174	-.302	.346
Q3.6	3	6	4.98	.799	-.330	.174	-.532	.346
Q3.7	4	6	5.21	.657	-.243	.174	-.728	.346
Q3.8	3	6	4.76	.738	-.128	.174	-.287	.346
Q3.9	3	6	5.06	.683	-.267	.174	-.182	.346
Q3.10	3	6	4.56	.725	.307	.174	-.373	.346
Q3.11	2	6	4.54	.747	-.001	.174	.104	.346
Q3.12	3	6	5.26	.639	-.403	.174	-.090	.346
Q3.13	3	6	5.20	.623	-.425	.174	.646	.346
Q3.14	2	6	4.83	.666	-.210	.174	.729	.346
Q3.15	1	6	4.33	.944	-.403	.174	.146	.346
Q3.16	2	6	4.59	.771	-.237	.174	.091	.346
Q3.17	3	6	4.51	.742	.012	.174	-.287	.346
Q3.18	2	6	4.49	.821	-.326	.174	.317	.346
Q3.19	3	6	4.92	.684	-.091	.174	-.358	.346
Q3.20	3	6	4.92	.676	-.001	.174	-.544	.346
Q3.21	3	6	4.94	.652	-.057	.174	-.312	.346
Q3.22	2	6	4.77	.862	-.620	.174	.353	.346
Q3.23	3	6	5.03	.673	-.344	.174	.222	.346
Q3.24	3	6	5.07	.677	-.691	.174	1.275	.346
Q3.25	3	6	5.04	.669	-.459	.174	.625	.346
Q3.26	4	6	5.24	.597	-.135	.174	-.474	.346
Q3.27	4	6	5.10	.631	-.077	.174	-.491	.346
Q3.28	4	6	5.21	.625	-.178	.174	-.562	.346
Q3.29	3	6	4.98	.626	-.113	.174	-.014	.346

Q3.30	3	6	4.96	.657	-.509	.174	.915	.346
Q3.31	3	6	5.13	.591	-.189	.174	.440	.346
Q3.32	4	6	5.07	.579	-.002	.174	-.015	.346
Q3.33	2	6	4.68	.832	-.811	.174	1.131	.346
Q3.34	2	6	4.71	.801	-.458	.174	.203	.346
Q3.35	3	6	4.96	.653	-.413	.174	.670	.346
Q3.36	2	6	4.63	.752	-.216	.174	.190	.346
Q3.37	3	6	5.04	.633	-.156	.174	-.047	.346
Q3.38	4	6	4.98	.605	.006	.174	-.228	.346
Q3.39	3	6	5.05	.615	-.164	.174	.152	.346
Q3.40	4	6	5.05	.612	-.024	.174	-.302	.346
Q3.41	3	6	5.06	.639	-.173	.174	-.106	.346
Q3.42	3	6	5.05	.620	-.160	.174	.097	.346

Appendix 8 Extract of Amy's Teaching Plan

The overall aims for a whole term (Grade 3, Term 2)

- *Will be curious about English and be willing to listen to and imitate others speaking English*
- *Can sing simple English songs and say simple English chants.*
- *Can understand and read simple stories with the help of pictures.*
- *Can play games, act out and do other activities (colouring pictures, matching pictures with words) according to teachers' instruction*
- *Can take active part in learning and cooperation*
- *Will be willing to know foreign culture.*

Learning objectives for a lesson of Unit 3 (At the Zoo)

- *Students can read and say new sentences (to describe animals): it has a....*
- *Students can read and say the words: small, big, long and short*
- *Students can use the sentences and words to introduce the animals and themselves correctly.*

