Implementing Classroom-Based Assessment for Young EFL Learners in the Chinese Context: A Case Study

Qiaozhen Yan¹ · Lawrence Jun Zhang² · Xiaolong Cheng³

Accepted: 11 July 2021 / Published online: 19 July 2021
© De La Salle University 2021

Abstract While there is extensive literature on how classroom-based assessment (CBA) can be effectively put into practice, little is known about its implementation in L2 contexts, especially in the young English-as-a-foreign-language (EFL) learner context. This study endeavored to investigate teachers' CBA practices and factors that might exert influences on them. A purposive sample of three EFL teachers from two primary schools participated in this case study. Our thematic analysis revealed that the potential of CBA in supporting young EFL learners' learning had not been fulfilled. The teachers failed to clarify such objectives and success criteria to their students. Despite the use of multiple assessments, the teachers relied heavily on formal assessments, with student-involving assessments being less frequently used. Moreover, there was a heavy reliance on norm-referenced assessment and evaluative feedback. It was also found that teachers' CBA practices faced complex challenges related to teacher, student, context and system factors. Practical implications

for how CBA can be effectively implemented in similar EFL contexts are discussed.

Keywords Classroom-based assessment · Young EFL learners · Teachers' practices · Influential factors

Introduction

Although classroom-based assessment (CBA) is increasingly recognized as crucial to language learning (Liu & Xu, 2017), research on L2 teachers' implementation of CBA is relatively scant. The limited body of research has focused on certain CBA aspects, such as using assessment methods and providing feedback (Chen et al., 2014; Cheng et al., 2008; Gan et al., 2018). A comparatively comprehensive understanding of L2 teachers' CBA practices is lacking, which is an important gap because CBA is a holistic concept (Lee et al., 2019). Moreover, most of the studies have been conducted in secondary and tertiary EFL contexts (Guo & Xu, 2020; Saito & Inoi, 2017), with little attention paid to the young EFL learner context. Young language learners (YLL), known as "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), are different from older or adult learners due to their unique cognitive, social, and emotional characteristics. Given that CBA has been advocated as an integral part of YLL teaching (Butler, 2019), more research is needed in this regard.

Informed by Davison and Leung's (2009) CBA framework, this study aimed to gain a comprehensive understanding of teachers' CBA practices in the young EFL learner context. Besides, the study examined how various

✉ Lawrence Jun Zhang
lj.zhang@auckland.ac.nz

Qiaozhen Yan
qiaozhencqu@163.com

Xiaolong Cheng
xl.cheng@auckland.ac.nz

¹ School of Foreign Languages and Cultures, Chongqing University, Chongqing, China

² Faculty of Education and Social Work, The University of Auckland, Private Bag 92601, Symonds Street, Auckland 1150, New Zealand

³ School of Foreign Languages, Hubei University of Technology, Wuhan, China

factors played a role in teachers' CBA practices. It is expected that such information could shed important light upon how CBA can be integrated into L2 contexts to support student learning, particularly under post-COVID-19 conditions. As the COVID-19 pandemic has led to some challenges in L2 education such as how to organize interactive activities and provide timely feedback (Gao & Zhang, 2020; Zhang et al., 2021), it is imperative to seek recommendations for facilitating student learning during this difficult period. The present case study, though conducted prior to the outbreak of COVID-19, examined EFL teachers' implementation of CBA in-depth. Thus, it is hoped that the information collected through this study can be of value for helping language teachers conduct CBA to promote student learning to tide over COVID-19.

Literature Review

Understanding Classroom-Based Assessment

The term CBA, used interchangeably with formative assessment, assessment for learning, and more recently learning-oriented assessment, generally refers to any assessment conducted by those who are directly responsible for teaching and learning, on an ongoing basis (Davison, 2019). From a technical perspective, CBA is viewed as a set of processes, encompassing collecting students' learning evidence, making judgments of the evidence, and using the evidence to make instructional decisions (Mcmillan, 2013). Traditionally conceived, CBA performs formative and summative functions, often understood as formative and summative assessment respectively. Impacted by an assessment for learning culture, however, CBA has been conceptualized as an important facilitator to the enhancement of student learning (Black & Wiliam, 2018). From this perspective, all assessments, including summative assessments, should be used formatively to facilitate student learning.

The defining features that make CBA formative are often described as three instructional processes: Identifying where learners need to go, where they are now in their learning and how to best get them there (Wiliam & Thompson, 2008). Central to the three processes is that learners should play an active role in assessment, such as setting learning goals, reflecting on and monitoring their learning, through which they develop self-regulation abilities to facilitate life-long learning (Elwood & Klenowski, 2010).

Informed by the key features of formative CBA, several frameworks have been developed to offer guidance on how to implement CBA to benefit student learning (e.g., Black & Wiliam, 2009; Davison & Leung, 2009; Ruiz-Primo,

2011). Nevertheless, there is no universally agreed framework in the existing literature, as Bennett (2011) aptly pointed out. Essentially, formative assessment is "both conceptually and practically still a work-in-progress" (p. 21). Davison and Leung's (2009) framework is regarded as the conceptual basis for the present study. It offers an operational means of examining teachers' assessment pedagogies by conceptualizing CBA as a cycle of four steps.

Specifically, planning assessment is the fundamental step, involving teachers clarifying learning goals and success criteria so that students understand where they are going (Timperley & Parr, 2009). While documents like national or state standards can be used as guidelines for goal setting (Abedi, 2010), students' learning needs also need to be taken into account to ensure that goals are attainable (Sadler, 1989). Teachers can clarify to students the learning goals using student-friendly language (Clarke et al., 2001), or engage students in discussing the learning goals (Ruiz-Primo, 2011).

Evidence collection, as a second step, can take diverse forms, often classified into three types: Spontaneous assessment opportunity, planned assessment opportunity, and formal assessment (Turner & Purpura, 2016). Spontaneous assessment opportunity, embedded in teacher-student interaction, can facilitate students to notice and understand the required standards. Planned assessment opportunity, which is also embedded in instruction but typically starts with students being engaged in an activity, can enable teachers to reflect and act on the evidence collected. Formal assessment, used at checkpoints, can allow teachers and students to know about their learning progress.

Subsequent to evidence collection, professional judgments are to be made. Criterion-referenced assessment can be conducted so that problems in learning and actions for improvement can be identified (Lok et al., 2016). Pupil-referenced assessment, which takes into account students' progress, also has a vital role to play in supporting learning (Harlen, 2006).

Finally, descriptive feedback is provided so that students can recognize the gap between their actual and designed learning goal and monitor their own learning to close the gap (Gamlem & Smith, 2013; Tunstall & Gipps, 1996). According to Tunstall and Gipps (1996), 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 of making choices for their own learning). Evaluative feedback, in contrast,

tells students whether they have performed a task correctly or not, providing little information about how to improve learning (Murtagh, 2014). This type of feedback often comes in the form of rewards (e.g., stickers, marks), punishments (e.g., removing rewards), expressions of approvals (e.g., smiling, general comment like “well done”), and disapprovals (e.g., criticizing students). It should be noted that, the four steps are interactively connected, contributing to the formative enactment of CBA collectively.

Research on Classroom-Based Assessment in L2 Contexts

The existing literature on CBA in L2 contexts has shown that the full potential of CBA has not been realized. For example, L2 teachers tend not to clarify learning goals and success criteria for students (Gu, 2014; Guo & Xu, 2020; Wu et al., 2021a,b). They employ multiple assessment methods, but with a major focus on traditional assessment methods, such as quizzes, multiple-choice items, rather than alternative assessments that foster a closer link between assessment and instruction, such as oral questioning, self-and peer assessment, and portfolios (Cheng & Sun, 2015; Gan et al., 2018). Previous studies have also revealed that L2 teachers mainly provide feedback in the form of grades and scores, while few efforts have been made to show students their strengths and weaknesses in learning (Xu & Liu, 2009; Zhou & Deneen, 2016). Furthermore, they fail to empower students to play an active role in language learning through self-assessment (Chen et al., 2014).

As prior research reveals, the implementation of CBA in L2 contexts is challenging, influenced by a number of contextual factors (Saito & Inoi, 2017; Wang et al., 2020). Drawing on Carless’ (2011) framework, these factors could be categorized into four levels: Teacher, student, school, and system. For instance, L2 teachers’ CBA practices are affected by teacher-related factors such as teachers’ assessment literacy and beliefs about learning and assessment (Gu, 2014; Xu & Liu, 2009), and student-related factors such as students’ attitudes towards learning and assessment (Lee et al., 2016). School-related factors include school assessment policy, heavy workload, class size, and support from administrators and colleagues (Cheng & Wang, 2007; Mak & Lee, 2014; Xu, 2016). At the system level, L2 teachers’ CBA practices are shaped by the examination-driven culture and the Confucian heritage culture (Chen et al., 2014; Cheng et al., 2008; Lee & Coniam, 2013).

Overall, two major gaps have emerged. First, the existing research on the implementation of CBA in L2 contexts seems to be fragmented as it has predominantly focused on some individual aspects of CBA, in particular, on how teachers

collect evidence and provide feedback (e.g., Cheng & Sun, 2015; Lee et al., 2016; Saito & Inoi, 2017). It leaves underexplored how teachers plan assessments and make professional judgments. Recently, Mak and Lee (2014) and Lee et al. (2019) have explored the implementation of CBA as a unitary concept in L2 writing classrooms in Hong Kong. However, prior to the investigation, the participating teachers from both studies had attended a workshop on assessment, held by one of the researchers to equip them with relevant knowledge of CBA. In Lee et al.’s (2019) study, continuous support was provided to teachers throughout the investigation. The two studies, therefore, fail to capture teachers’ assessment practices in a naturalistic setting. By offering a holistic picture of contemporary L2 teachers’ CBA practices, it is hoped that useful insights will be provided for L2 researchers and practitioners with regard to how to realize the full potential of CBA. Second, the majority of prior studies were conducted in secondary school or tertiary EFL contexts (e.g., Chen et al., 2014; Guo & Xu, 2020; Lee et al., 2016), with little attention given to CBA in the young language learner context. Different from adult learners, young language learners are undergoing cognitive, social, and emotional growth, and are vulnerable to adults’ praise and criticism (Mckay, 2006). These unique characteristics place demands on the assessment of young learners. Recent research has suggested that CBA offers great promise in promoting YLL learning as it elicits what children know and provides formative feedback, which can enhance motivation and foster children to become life-long self-regulated learners (Butler, 2019). While previous research has provided insights into how secondary school and university EFL teachers implement CBA practices, whether primary school teachers manifest similar assessment practices remains unclear. Extending this line of research will add to our limited knowledge of CBA for young language learners, shedding light upon how to effectively use CBA to promote YLL learning. Against this background, this study aims to investigate the implementation of CBA for young EFL learners in the Chinese primary school context. Specifically, it seeks to address the following research questions:

- (1) How do primary school EFL teachers implement CBA practices?
- (2) What are the factors influencing primary school EFL teachers’ CBA practices?

Methodology

Context and Participants

The Chinese Ministry of Education (MOE) has undertaken an important initiative to incorporate formative CBA into

the original summative assessment system of China. At the primary school level, an array of CBA principles are advocated in the New English Curriculum Standards for Compulsory Education (2011 version) (MOE, 2011), including establishing learning objectives and success criteria, using multiple assessment methods and engaging students in the assessment process.

The participants were recruited using maximum variation sampling (Merriam, 2016), which aimed to select participants representing a wide range of variations in demographics (e.g., teachers who differ in age, grade level taught, educational qualification, and teaching experience). This enables us to identify important shared patterns that cut across the sampled diversity. Three teachers from two public primary schools in a southwestern Chinese city were recruited. Amy (pseudonym) worked in School A (pseudonym), whereas Doris and Kathy worked in School B. The two schools differed in three ways. School A had a large school size, with around 70 classes, which doubled the size of School B. It recruited a larger number of high quality teachers who were graduated from top teacher education universities, where they had acquired new theories of teaching. Regular research activities were also organized in School A to discuss how to apply theories in practice to improve teaching. Table 1 provides an overview of the three teachers' background information.

Data Collection

In this study, multiple sources of data were collected, including classroom observations, semi-structured interviews, and documents.

Classroom Observations

Classroom observations were conducted with each teacher four times (40 min for each) over a period of six weeks during a term. The foci of the observations were on teachers' CBA practices with regard to planning assessment, collecting learning evidence, making judgments, and providing feedback. The first researcher conducted the observations as a non-participant observer, during which all lessons were audio recorded and field notes were taken.

Semi-Structured Interviews

One semi-structured interview (approximately one hour) was conducted with each teacher within two days after the final observation. The interview questions covered teachers' background information, CBA practices, and difficulties in implementing CBA. The interviews provided a main source of data about factors in teachers' CBA practices and triangulated with the observational data about teachers' practices. The time and place for interviews were selected by teachers for their convenience. All interviews were audio recorded and conducted in mandarin Chinese to ensure that they were at ease. Notes were taken during each interview.

Documents

Relevant documents were collected to provide supplementary information about teachers' CBA practices, including copies of lesson plans, curriculum standards, textbooks, and teaching slides.

Data Analysis

The audio recordings of 12 observed lessons and three interviews were transcribed. The interview transcripts were not translated into English, as the difference between the source and target language can cause loss of information. Then, all the qualitative data, including documents, were discerned and sifted using thematic analysis (Miles et al., 2014).

To address RQ1, the observation and interview transcripts were analyzed to identify patterns in teachers' CBA practices. The transcripts were repeatedly read and carefully examined, leading to a range of codes regarding teachers' CBA practices. The codes from the dataset of each teacher were compared to generate the main themes. Such themes were grouped into four major categories, based on Davison and Leung's (2009) framework. Relevant documents were also reviewed accordingly. To address RQ2, the interview transcripts were analyzed, using a similar analysis approach. The data were carefully processed and coded to identify emerging themes, which were compared and reduced into overarching categories according to Carless' (2011) framework.

Table 1 Background information on three teachers

Names	Gender	Age	Educational qualification	Major	Teaching experience (years)	Grade level taught
Amy	Female	24	Bachelor	Primary Education (English)	2	3, 4, 6
Doris	Female	29	Master	English Language Teaching	4	4
Kathy	Female	40	Bachelor	English Language Teaching	12	6

Findings

Teachers' CBA Practices

The teachers' CBA practices emerged in four dimensions: Planning assessment, collecting learning evidence, making professional judgments, and providing feedback.

Planning Assessment

According to the document data, all teachers attempted to unpack the national curriculum standards into specific instructional objectives, which were categorized into three levels: Term, unit, and lesson. The term-level objectives were general in nature, established according to the national curriculum standards; the unit-and lesson-level objectives gave a more detailed description of language skills, knowledge, and affective attitudes that students were expected to acquire. The teachers stated that the national curriculum standards enabled them to have a big idea of where students were going. Kathy, for instance, recorded as follows:

The national curriculum standards were a set of unifying standards for primary school students, specifying the overall requirements that students at Grade 1, 2 and 3 were expected to meet.

Kathy admitted that she adjusted instructional objectives according to students' language proficiency. The overall requirements specified by the national curriculum standards were relatively unachievable for her students. She explained:

Students from less developed cities were with a relatively low language proficiency. It was difficult for them to meet the national curriculum requirements.

Doris and Amy also confessed that they established vocabulary learning objectives on the basis of textbooks. According to Amy, "*Textbooks offer a description of vocabulary requirements.*" However, the teachers failed to explicitly clarify learning objectives. They all followed a similar instructional procedure (namely lead-in, language point explanation, and check-out), focusing predominantly on students' task performance and providing little scaffolding to help students understand the required learning objectives. Consequently, students were unclear about where they were going. For instance, in one lesson, Amy used a riddle about animals as a lead-in of a spelling unit, following which she taught students how to spell the letter "I" with a vowel/i:/and checked their attainment. During the whole process, she provided no opportunity to enable students to understand what they were expected to learn.

As a result, most students had difficulty in recognizing the learning objective by the end of the lesson, stating that they had learned the words "*Big*" and "*Pig*" rather than the spelling of letter "I".

Besides, the teachers drew students' attention to the process of completing an assessment task but did not articulate success criteria to students. In Amy's class, for instance, she familiarized students with the steps of completing a recitation task. However, she did not explain to students what constituted successful recitation performance.

Collecting Learning Evidence

The observational data show that the teachers used three major types of assessments to elicit learning evidence. First, they frequently used spontaneous assessment opportunities, including questioning and observation. Closed questions were mainly used to acquire factual information, or simple comprehension, from individual students or the whole class. Such questioning practice enabled teachers to modify their instruction if required. In one observed lesson, Doris was teaching a weather report played on the radio, during which she asked several questions, such as "*It will be?*" and "*What else?*," to see whether students could understand the content of the report. Based on students' responses, she modified her teaching, asking students to listen to the radio again to ensure their understanding of the main information.

The teachers also consistently observed students' reactions, such as reading in a low voice and keeping silent in response to teachers' questions. This enabled the teachers to know how students were progressing towards the expected objective and to make further instructional decisions. For example, in one of Kathy's lessons, while students were engaged in a group discussion to describe a picture using the sentence pattern "*What's happening?*," Kathy noticed that some students asked about the meaning of the word "*Everything*," and she explained this word to the whole class.

Second, the teachers heavily relied on formal assessments, including oral-reading tasks and textbook exercises. Typically, after new words and sentences had been taught, students were asked to complete these formal assessment tasks. They mainly served a summative purpose as evaluative feedback such as scores and general praises were provided to check whether students had mastered the target linguistic knowledge. For example, Doris commented on students' oral-reading performance by offering general praise "*Good job.*" Kathy asked students to complete a textbook exercise and clearly explained how each item was scored. She also constructed exercises, such as multiple-choice items, to check students' attainment. Kathy

commented that such exercises were used to cultivate students' test-taking skills.

Planned student-involving assessment opportunities, such self-and peer assessment, were not a regular practice for Kathy. She occasionally asked students to make summative judgments of their own performance by the end of a month in the form of grade levels "A, B, C." As for Doris, while students were engaged in peer assessment, general praises such as "Good" or "Wonderful" were the major feedback form, providing little information regarding how to improve learning.

Amy was the only teacher who attempted to use self-and peer assessment to locate difficulties in students' learning. However, she did not make explicit success criteria, which resulted in students being unclear about what constituted the quality of good work. As a consequence, they tended to evaluate their own or peers' performance in a summative way. For instance, when one student was asked to evaluate his own handwriting, he could only make a summative judgment of his overall performance by stating that "*I have not done a good job.*" In another example, although peers were asked to offer feedback on the strengths and weaknesses of one student's oral presentation, they tended to ignore the strengths and focus on the weaknesses, by saying that "*He spoke too fast*" and "*He did not show any emotion.*"

Making Professional Judgments

When making judgments, the teachers seldom set specific criteria against which students had to work. They occasionally used students' prior performance against which to interpret assessment information. For instance, Amy pointed out students' progress in handwriting one letter. The teachers all believed that it was crucial that children recognize their learning progress and experience success. Through pupil-referenced assessment, children were given opportunities to experience such kind of success, thus, enhancing their motivation. As indicated by Kathy:

By comparing students' current and past achievement, we could point out the progress made by students. This encouraged them to make further progress.

In contrast, norm-referenced assessment was a common practice. The teachers frequently made comparisons between groups or individual students, using general praises or giving stickers to those who had achieved a good level of performance. For example, in one oral-reading task, Doris asked students to compete in groups to win stickers. However, through making comparisons, a competitive classroom environment was created, which exerted a negative impact on students' social relationships and self-

confidence. For instance, in Doris' class, while students who won stickers were cheering for their success in an oral-reading task, those who lost the stickers became disappointed and upset.

The teachers were concerned about the negative impact of norm-referenced assessment on students' self-esteem and felt the need to properly use such assessment. Doris said:

I thought this would impact students' self-esteem, so teachers should use it carefully.

However, all three teachers admitted that by comparing students' summative performance, lower achieving students could be motivated to achieve a higher achievement. Kathy explained as follows:

Students could easily recognize the difference when asked to compare their scores with each other, thus being motivated to work hard.

Providing Feedback

The teachers occasionally provided descriptive feedback to specify improvement, helping students to identify and correct errors. This type of feedback was provided in two major forms. First, the teachers pointed out directly what was wrong and used oral questions to guide students to correct errors. In one lesson, Amy observed that students had made a mistake in pronouncing the word "*Banana*". She immediately pointed out the error and asked a question "*How to spell the first letter A?*" to help students correct the pronunciation. Second, students were sometimes asked to self-check their errors. For example, while students were doing a textbook written exercise, Amy asked them to check their handwriting to see whether they had written the letter "*K*" correctly, as they had confused the capital letter *K* with lowercase letter *k* in the previous lesson.

In contrast, the qualitative data show a heavy reliance on evaluative feedback, which was categorized into four types: Rewarding, punishing, approving, and disapproving. In terms of rewarding, the teachers frequently used symbols, such as stickers and stars, to motivate students who had made efforts in their work or behavior. In one example, Amy gave a sticker to one student who had tried to correct the pronunciation of the word "*Six*." Bonus scores, another notable form of rewarding, were used to make a summative evaluation of students' work. For instance, in one lesson, Amy asked students to spell together the letters they had learnt in a previous lesson. Each group was awarded one bonus point as "*Every student in the group has done a good job.*" Oppositely, punishing was used to express dissatisfaction with students who were not concentrated on their learning or had not made any effort in their work. In

Doris' class, she removed one sticker from one group, as a punishment, as they were not following her instruction carefully. Approving, an overall expression of satisfaction was used when the teachers judged that students' learning attitudes and achievement were satisfactory. It often took the form of general praises like "*Very good*," "*Well done*," and "*Good job*." Disapproval was expressed when teachers judged that the lack of effort or concentration was the cause of students' poor performance. In one lesson, Amy became disappointed with those students who had not completed a recitation task at home, saying "*Have you recited the talk at home?*" in a loud voice. The teachers all believed that evaluative feedback was helpful for motivating students to achieve the expected level of performance. Amy, for example, noted:

The stickers or stars students would be transferred into prizes at the end of the term. Thus, every student would be motivated to work hard to win a prize.

To sum up, the findings indicate that despite teachers' attempt to establish clear instructional objectives, they failed to clarify the learning objectives to students. They relied heavily on formal assessments for summative evaluation, whereas student-involving assessments were less frequently used and its potential in helping students self-monitor their learning was not realized. They used norm-referenced assessment consistently as the basis of making judgments and frequently provided evaluative feedback to motivate students.

Factors Influencing Teachers' CBA Practices

The findings show that teachers' CBA practices were influenced by a variety of factors related to four levels: Teacher, student, school and system.

Teacher-Related Factors

The teachers were not equipped with professional knowledge and skills about CBA, which was associated with inadequate assessment literacy training. Although Doris and Amy had attained a course on assessment during their master's and bachelor's programs, respectively, CBA was not a major component. Instead, the course focused mainly on large-scale language tests. For example, Doris said:

Our course instructor focused on analyzing examination papers for senior secondary school students. Little attention was given to CBA.

Both Amy and Kathy had received no CBA training during in-service teacher education programs. Doris, the only teacher who had such experience, confessed that the content was theory driven, providing little practical

guidance. The teachers all felt a need for in-service teacher education on CBA so that they could be equipped with essential skills. As Doris said:

In-service training in CBA was beneficial to our assessment practices. I would be very much willing to receive such kind of training.

Another teacher-related factor was teachers' beliefs about scores. The teachers firmly believed that scores played an important role in enhancing students' motivation, which posed a challenge to the effective implementation of CBA. For instance, while the teachers admitted that norm-referenced assessment had a negative impact on students' self-esteem, they held strong beliefs that by comparing students' summative performance, lower achieving students could be motivated to make efforts to achieve better performance. This explained why the teachers relied heavily on norm-referenced assessment.

Such strong beliefs about scores also contributed to teachers' regular provision of evaluative feedback. According to Amy, the use of bonus scores encouraged students to work hard in order to win a prize at the end of a term. Similarly, Kathy believed that giving bonus scores was encouraging to both students themselves and their peers:

When this student got bonus scores both the student himself and his peers would be motivated to do a better job next time.

Student-Related Factors

Students' learning needs posed a challenge to the clarification of learning objectives. The teachers stated that primary school students, as young learners, often lacked interest or had difficulty in understanding the learning objectives. Doris, for instance, used to begin her lesson with an explanation of the learning objectives but found that students "*showed no interest*" in knowing what they were expected to learn. As a consequence, she decided not to make the learning objectives explicit to students. Instead, she designed a topic-centered unit, hoping that students could have a sense of what they were expected to learn by the end of the unit by themselves. She reflected as follows:

I presented the lesson objectives, but they were not interested in knowing them.

Kathy admitted that, as for her students who were from Grade 6 (around 12 years old), it was appropriate to explain the learning objectives at the beginning of a lesson; however, as for younger children, they often had difficulty in understanding such learning objectives:

If you explained what they were going to achieve at the very beginning, they would get puzzled, wondering what it meant.

School-Related Factors

In the study, Amy and Kathy were teaching four classes and Doris had six classes. The class size was large, with an average number of 40 to 50 students in each class. Therefore, certain CBA practices were just occasionally used. In Kathy's case, the large classes had constrained the provision of detailed feedback to individual students. Amy explained that large class size hindered her from consistently observing their students' reactions.

Grade level also shaped teachers' CBA practices. All three teachers relied on formal assessments such as textbook exercises. However, a major difference was that Kathy teaching at a higher grade level (Grade 6) was the only teacher who regularly used teacher-constructed exercises to measure students' outcome. As Kathy explained, her students were about to graduate from primary school and to enroll into junior secondary school the next semester. Exams would then become a common practice and summative scores inevitably played a pivotal role in their life. Therefore, she designed teacher-constructed exercises and evaluated students' performance with scores, with a view to preparing students for high-stakes exams.

Besides, the teachers were obligated to complete a packed syllabus, which became another constraint. In Doris' case, she started her classes from 10 a.m. in the morning and did not finish classes till afternoon every workday in order to complete a whole textbook within one semester. Because of the heavy curriculum workload and the time constraint, she was unable to devote extra time and efforts to design assessment activities, such as self-and peer assessment. She felt the need to have a flexible syllabus, as she commented:

The major reason was that I had a tight teaching schedule. Only if I had sufficient time could I implement self-and peer assessment more effectively.

The lack of collaborative support from colleagues was another challenge. Doris teaching in School B mentioned that her colleagues would work collaboratively to help solve pedagogical problems only when she needed to participate in a pedagogy competition. During daily teaching, however, she had to conduct assessment practices on an individual basis. Different from Doris, Amy teaching in School A stated that teaching and research meetings were held each week in her school, in which colleagues worked together to discuss and solve problems in their classroom practices. This enabled their school to have a

foundation for implementing new curriculum ideas. This partly explained why Amy was the only teacher among the three who attempted to use self-and peer assessment to identify students' learning gaps.

System-Related Factor

At the system level, teachers' CBA practices were influenced by the examination-driven culture in Chinese schools. The document data show that the assessment policy of two schools put a heavy emphasis on summative scores. The assessment structure was composed of two parts: On-going assessment (e.g., students' in-class performance and assignments) and final exams. Students' performance in both parts was evaluated with scores, which resulted in the summative nature of assessment policy. Although teachers attempted to implement some formative CBA activities, such as self-and peer assessment and descriptive feedback, they needed to follow the assessment policy. This, therefore, constituted an impediment to the implementation of CBA.

Discussion

The aim of the study was to gain a comprehensive understanding of primary school EFL teachers' CBA practices and the influencing factors in their assessment practices. The discussion will predominantly focus on the two perspectives.

Teachers' CBA Practices

The findings show that the teachers in the study did not fully utilize CBA to facilitate young EFL learners' learning, lending support to a worldwide concern that teachers' CBA practices remain relatively weak (William, 2010). The teachers attempted to establish appropriate instructional objectives that reflected national curriculum standards, students' language proficiency, and the textbooks being used, which could enable them to plan instructional activities and select appropriate assessments (Ruiz-Primo, 2011). However, they did not explain explicitly to students the learning objectives and success criteria. This finding is in line with that of Zhou and Deneen (2016), though it differs from that of Lee et al. (2019), which found that Hong Kong primary writing teachers put a great emphasis on sharing learning objectives and success criteria. A possible reason for the difference is that the participating teachers of Lee et al. (2019) attended a teacher workshop on assessment as learning.

Consistent with prior research (Gan et al., 2018), this study found that the teachers used a wide range of

assessment methods, including spontaneous assessments, planned student-involving assessment opportunities, and formal assessments. A possible explanation could be that multiple assessments have the potential to cater to the varied learning needs of language learners (Leung, 2005). Among the multiple assessment methods, student-involving assessment opportunities, such as self-and peer assessment, were the least frequently used. This finding, as reported in previous studies (Chen et al., 2014; Saito & Inoi, 2017), may be explained by the Confucian heritage culture in China, which is featured by a hierarchical student–teacher relationship (Carless, 2011).

The teachers' practices of making judgments did not seem to support student learning, as criterion-referenced assessment was not used consistently as the basis of judgments (Harlen, 2007). The occasional use of pupil-referenced assessment provides evidence that young learners' age-related characteristics, being sensitive to success or failure, places demands on teachers' assessment use (Butler, 2019). By contrast, the teachers regularly used norm-referenced assessment, putting an emphasis on comparing students' academic results. This might be impacted by the dominant culture of standardized testing in the Chinese EFL teaching context, where scores play a pivotal role in students' life (Cheng, 2008).

The teachers attempted to provide descriptive feedback, such as explicitly pointing out students' errors, using oral questions to guide students to correct their errors and involving students in self-check of their errors. The first two can also be referred to as explicit correction and elicitation, which are two major types of oral corrective feedback (Lyster & Ranta, 1997). However, they put more emphasis on evaluative feedback. This finding is in contrast with that of Chen et al.'s (2014) study, showing that two Chinese university EFL teachers provided specific and supportive feedback. Overall, this study lends empirical support to Guo and Xu's (2020) and Zhou and Deneen's (2016) findings that Chinese EFL teachers relied heavily on evaluative feedback. This reliance on evaluative feedback might also be explained by the high-stakes testing culture of China.

A synthesis of the findings suggests that primary school EFL teachers did not seem to make full use of CBA to facilitate YLL learning. This lends support to Butler's (2019) concern that teachers' practices of CBA for young learners remain weak. This finding is also consistent with previous research suggesting that potential of CBA has not been realized in the secondary school and tertiary EFL contexts (e.g., Chen et al., 2014; Guo & Xu, 2020).

Factors Influencing Teachers' CBA Practices

The study found that the implementation of CBA in Chinese primary school EFL context was influenced by a range of factors, related to teacher, student, context, and system, which echoes the factor framework of Carless (2011). In general, this finding indicates that the implementation of CBA is challenging and highly contextualized (Davison, 2019).

First, the teachers were found to lack essential CBA knowledge and skills due to insufficient professional training. The finding on the lack of CBA literacy training for EFL teachers has been commonly reported in previous research (Xu, 2017; Xu & Liu, 2009). The study also revealed the necessity of professional training on CBA for in-service EFL teachers, echoing Lan and Fan's (2019) study. This finding suggests that language teachers' CBA literacy is underdeveloped (Lam, 2015).

The teachers also believed that scores played a role in enhancing students' motivation, thus, relying heavily on norm-referenced assessment and evaluative feedback. This finding provides evidence to show that teachers' core beliefs powerfully influence their classroom practices (Phips & Borg, 2009; Sun & Zhang, 2021). Teachers' core beliefs are stable and highly resistant to change and, thus, can have profound impact on teachers' classroom practices. In the study, while the teachers were concerned about the negative impact of norm-referenced assessment on students' self-esteem, they strongly believed that comparisons of students' summative scores allowed students to be greatly motivated to achieve better performance. Such strong beliefs can be regarded as deep-rooted and core beliefs, which might be shaped China's well note examination culture, and thus resulted in teachers' regular use of norm-referenced assessment.

At the student level, due to the fact that young learners had difficulty or lacked interest in understanding the learning objectives, the teachers tended not to clarify these objectives. This reinforces Butler's (2019) argument that it is important to consider the cognitive demands that YLL assessment tasks entail and the degree of interest that young learners have in the tasks.

At the school level, large class size provided unfavorable conditions to the provision of descriptive feedback. Indeed, large class size has been widely acknowledged as an obstacle for the effective implementation of CBA in L2 contexts (Xu, 2016). The results also suggest that teachers teaching at a higher grade level were more likely to use formal assessments like teacher-constructed exercises. Similarly, Zhang and Burry-Stock's (2003) study showed that as the grade level increased, teachers tended to rely on objective formal assessment techniques. Heavy workload also posed a major impediment to the full uptake of CBA,

which is consistent with previous research (Mak & Lee, 2014). Support from colleagues constituted another challenge, agreeing with that of Lee et al. (2016).

Finally, the study demonstrated that the examination-driven culture of China emerged as significant in influencing teachers' CBA practices. L2 education in China is characterized by an examination-driven culture, in which scores are considered as the key to success (Cheng, 2008). Thus, teachers are likely to focus on summative scores, paying less attention to the strengths and weaknesses in student learning.

Conclusion

This study investigated the implementation of CBA in the Chinese young EFL learner context. We found that the teachers did not fully implement CBA practices to support young EFL learners' learning. From a social culture perspective, the implementation of CBA was constrained by multiple challenges that stemmed from key stakeholders (teacher and student) and external school and social contexts.

This study has several implications for implementing CBA in similar young EFL learner contexts. As mentioned previously, it is hoped these implications can be valuable for CBA implementation in the post-Covid-19 era. First, the study shows that language teachers' CBA literacy is underdeveloped. Future teacher education programs can include content knowledge about CBA and offer related practical guidelines to equip teachers with necessary CBA expertise. As noted by Zhang et al. (2021), the Covid-19 pandemic imposes the challenge in using technology for assessment. Therefore, professional training can be offered to improve teachers' expertise in incorporating technology in CBA. Second, given the workload influence on teachers' CBA use, school administrators can grant teachers autonomy in planning their teaching schedule, enabling teachers to put new assessment ideas into practice. The study also provides evidence about the influence of collaborative teaching environment. Therefore, school administrators can find it useful and professionally rewarding if they foster a collaborative professional community. Meetings and discussions can be held to convey the principles of CBA to all staff; colleagues can be encouraged to share assessment experiences. Third, since young learners' characteristics exert an impact on teachers' assessment practices, primary school teachers should consider young learners' needs when carrying out CBA activities. Overall, this study demonstrates that CBA implementation is challenging and contextualized. Teachers need to raise their awareness of various influential factors and be ready to deal with potential challenges that might arise from the learning and

teaching process. This implication is particularly significant for teachers in the post Covid-19 era as they are likely to face unexpected challenges (Zhang et al., 2021).

However, the present study is not without limitations. First, only three teachers from one city were recruited, accounting for a small sample. Thus, caution should be exercised when the relevance of the findings is considered in relation to teachers' CBA practices in other contexts. Second, this study investigated CBA implementation from the perspective of the language teachers themselves. Students, as an important stakeholder in the educational system, may hold different views. Future studies that incorporate the views of both teachers and students can provide further insights into how to better reap the benefits of CBA.

Author Contributions QY collected the data and wrote the first draft. LJZ contributed to the conceptualisation of the research design, helped with the data analysis, and revised the subsequent versions; XC helped with the revision of the subsequent version; LJZ further revised and finalized the manuscript for submission as the corresponding author.

Funding The work was partially supported by the Fundamental Research Funds for the Central Universities, China (Grant number: 2021CDJSKZX07).

Data Availability The study reported in this paper is a qualitative study, where the original data can be provided, but such data are also used to support the first author's Ph.D. Thesis.

Code Availability The data were analyzed firstly by using Endnote and then manually. A coding sample can be provided.

Declarations

Conflict of interest There are no financial and non-financial competing interests in relation to the manuscript.

References

- Abedi, J. (2010). Formative assessment with English language learners. In H. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 181–197). Routledge.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25.
- 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 & Practice*, 25(6), 551–575.
- 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). Springer Nature.

- Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. Routledge.
- 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 & 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., Rogers, T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & 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., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, 4(1), 85–107.
- Clarke, S., Timperley, H., & Hattie, J. (2001). *Unloacking formative assessment: Practical strategies for enhancing students' learning in the primary and intermediate classroom*. Hodder and Stoughton.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415.
- 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.
- Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & 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. (2020). Teacher learning in difficult times: Examining foreign language teachers' cognitions about online teaching to tide over COVID-19. *Frontiers in Psychology*, 11(549653), 1–14.
- Gu, Y. (2014). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a foreign language in a Chinese secondary school? *Assessment in Education: Principles, Policy & Practice*, 21(3), 286–305.
- Guo, Q., & Xu, Y. (2020). Formative assessment use in university EFL writing instruction: A survey report from China. *Asia Pacific Journal of Education*. <https://doi.org/10.1080/02188791.2020.1798737>
- Harlen, W. (2006). *Helping learning: A framework for decisions* (4th ed.). Paul Chapman.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197.
- Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, 61, 112–122.
- 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.
- Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 526–537). Lawrence Erlbaum Associates.
- Liu, J., & Xu, Y. (2017). Assessment for learning in English language classrooms in China: Contexts, problems, and solutions. In H. Reinders, D. Nunan, & B. Zou (Eds.), *Innovation in language learning and teaching: The case of China* (pp. 17–37). Palgrave McMillan.
- 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.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37–66.
- Mak, P., & Lee, I. (2014). Implementing assessment for learning in L2 writing: An activity theory perspective. *System*, 47, 73–87.
- McKay, P. (2006). *Assessing young language learners*. Cambridge University Press.
- McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 2–17). Sage.
- Merriam, S. B. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.
- MOE. (2011). *English curriculum standards for compulsory education (2011 version)*. Beijing Normal University Press.
- Murtagh, L. (2014). The motivational paradox of feedback: Teacher and student perceptions. *The Curriculum Journal*, 25(4), 516–541.
- Phipps, S., & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System*, 37, 380–390.
- Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37, 15–24.
- 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.
- Sun, Q., & Zhang, L. J. (2021). A sociocultural perspective on English-as-a-foreign-language (EFL) teachers' cognitions about form-focused instruction. *Frontiers in Psychology*, 12(593172), 1–15.
- 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.
- Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, 22(4), 389–404.
- Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 1–19). De Gruyter Mouton.
- Wang, L., Lee, I., & Park, M. (2020). Chinese university EFL teachers' beliefs and practices of classroom writing assessment. *Studies in Educational Evaluation*, 66, 100890.
- Wiliam, D. (2010). 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). Routledge.
- Wiliam, 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). Lawrence Erlbaum Associates.
- Wu, X. M., Zhang, L. J., & Dixon, H. R. (2021a) Implementing Assessment for Learning (AfL) in Chinese university EFL classes: Teachers' values and practices. *System*, 102589, 1–12. <https://doi.org/10.1016/j.system.2021.102589>
- Wu, X. M., Dixon, H. R., & Zhang, L. J. (2021b). Sustainable development of students' learning capabilities: The case of university students' attitudes towards teachers, peers, and themselves as oral feedback sources in learning English. *Sustainability*, 13(9), 5211, 1–12. <https://doi.org/10.3390/su13095211>
- Xu, H. (2017). Exploring novice EFL teachers' classroom assessment literacy development: A three-year longitudinal study. *Asia-Pacific Education Researcher*, 26(3–4), 219–226.
- 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., & 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.
- Zhang, C., Yan, X., & Wang, J. (2021). EFL Teachers' online assessment practices during the COVID-19 pandemic: Changes and mediating factors. *Asia-Pacific Education Researcher*. <https://doi.org/10.1007/s40299-021-00589-3>
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323–342.
- 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.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.