

*Chinese secondary school teachers' conceptions of
L2 assessment: A mixed-methods study*

Maggie Ma

The Hang Seng University of Hong Kong, China

<https://orcid.org/0000-0002-9805-5100>

maggiema@hsu.edu.hk

Gavin Bui

The Hang Seng University of Hong Kong, China

<https://orcid.org/0000-0002-1567-9074>

gavinbui@hsu.edu.hk

Abstract

Teacher conceptions of assessment influence their implementation of learning-focused assessment initiatives as advocated in many educational policy documents. This mixed-methods study investigated Chinese secondary school teachers' conceptions of L2 assessment in the context of an exam-oriented educational system which emphasizes English grammar, vocabulary and reading comprehension skills. For the quantitative part of the study, survey data were collected to gauge the conceptions of assessment held by 66 senior secondary EFL teachers from six schools in Eastern China. For the qualitative part, case studies of two teachers from schools with different rankings were conducted. Quantitative results showed that the teacher participants as a group agreed most with the view that assessment is to help learning. However, there was a strong association between two factors, that is, the assessment as accurate for examination and teacher/school control factor, and the assessment as accurate for student development factor. The strong association indicated that it may be less likely for the group of teachers to adopt the formative assessment initiatives emphasizing student development as promoted in the English curriculum reform. Qualitative findings further revealed individual differences in the two case study teachers' conceptions and practices of

assessment as well as the interplay among meso-level (e.g., school factor), micro-level (e.g., student factor), and macro-level (e.g., sociocultural and policy contexts) factors in shaping the teachers' different conceptions and practices of assessment. A situated approach has been proposed to enhance teachers' assessment literacy.

Keywords: Chinese EFL teachers; teachers' conceptions of assessment; assessment practices

1. Introduction

Assessment plays an important role in affecting students' learning. In recent years, many countries, including China, have witnessed the promotion of formative assessment (Berry & Adamson, 2011; Kennedy & Lee, 2008), which originated from England in response to the negative influence of high-stakes national testing (Stobart, 2006). The success of assessment innovation such as formative assessment relies much on teachers, who are the key agents in educational assessment (Xu & Brown, 2016). In particular, teacher beliefs regarding assessment may influence how they respond to learning-focused assessment and the success of its implementation (Brown et al., 2011). A lack of teacher beliefs in the proposed assessment innovation may constitute an obstacle to its success and calls for extensive assessment training. In countries where there is an exam-oriented educational system, it is thus crucial to understand teachers' views of assessment both for the success of policy initiatives and teachers' professional development.

This paper explores Chinese secondary EFL teachers' conceptions of assessment, defined as "a teachers' understanding of the nature and purpose of how students' learning is examined, tested, evaluated or assessed" (Brown & Gao, 2015, p.4). This is because teacher conceptions exert a major influence on how teachers perceive, respond to and interact with their teaching environment (Marton, 1981). Acknowledging that teacher conceptions of assessment are ecologically rational, previous research has investigated these conceptions in different contexts and resorted to macro-level factors (i.e., social and cultural factors) for an explanation (e.g., Brown et al., 2011; Brown & Michaelides, 2011; Teng & Bui, 2020). Despite such research, there is limited research on the influence of meso-level (e.g., school factors) and micro-level (e.g., characteristics of individual teachers) factors on teacher conceptions of assessment and their interplay with macro-level factors, particularly in the case of nationally advocated formative assessment innovation in exam-oriented educational contexts. Given that different levels of factors may shape teachers' assessment knowledge, beliefs,

and practices (Fulmer et al., 2015), it is important to explore how these factors affect teacher conceptions of assessment to shed light on the successful implementation of formative assessment and assessment training. To address the research gap, this study adopted a mixed-methods approach to examining Chinese secondary EFL teachers' conceptions of assessment and different layers of factors that shaped such conceptions when the recent English language curriculum reform has foregrounded the importance of formative assessment in the context of an exam-oriented educational system, which emphasizes English grammar, vocabulary and reading comprehension skills (Hao & Otani, 2016). The findings of the research may provide insights into the facilitation of the implementation of English education assessment initiatives and EFL teachers' professional development.

2. Literature review

2.1. Teachers' conceptions of assessment

Teachers hold beliefs about particular things (Pajares, 1992) and use their beliefs to filter new information, frame problem spaces, and guide actions (Fives & Buehl, 2012). In the context of assessment, teachers' beliefs about the nature and purposes of assessment, that is, their conceptions of assessment, may influence their assessment practices and create a lens through which they respond to curriculum and assessment reforms. For example, in societies with an exam-oriented educational system, teachers may hold the belief that a powerful way to improve student learning is to examine them, and they may be less likely to adopt formative assessment initiatives in educational reforms (Brown et al., 2011).

Research on teachers' conceptions of assessment, conducted extensively by Brown and his colleagues, has identified four major purposes of assessment based on the *Teacher Conceptions of Assessment* (TcoA) inventory (Brown, 2004, 2011; Brown et al., 2011; Brown & Michaelides, 2011). These purposes include: (1) assessment as improvement of teaching and learning (improvement); (2) assessment as making schools and teachers accountable for their effectiveness (school accountability); (3) assessment as making students accountable for their learning (student accountability); and (4) assessment as fundamentally irrelevant to the work and life of teachers and students (irrelevance). The first three are categorized as "purposes" while the last one is termed an "anti-purpose." When the school and student accountability views of assessment are grouped together, it seems that there are two major purposes of assessment in society, that is, accountability and improvement (Brown & Gao, 2015). This illustrates the dual functions of assessment and the potential tension that may arise from these two functions (Brown et al., 2011). On the one hand, assessments are utilized to

evaluate the effectiveness of teachers and schools and to certify the learning of students (i.e., the measuring and evaluative functions of assessment), but on the other hand, assessments are employed to inform different stakeholders (e.g., parents, teachers, students, governments, administrators) of learning progress and to enhance teaching and learning (i.e., the formative function of assessment).

Survey research using the TcoA has been conducted to explore teacher conceptions of assessment. Teachers strongly endorsed the notion of using assessment to improve teaching and learning. For example, secondary school teachers in New Zealand and teachers in Cyprus agreed most strongly with the view that assessment is used to improve learning (Brown, 2011; Brown & Michaelides, 2011). While they still agreed with using assessment to evaluate students, they viewed assessment as evaluating schools in a relatively negative light (Brown, 2011; Brown & Michaelides, 2011). Teachers rejected the conception that assessment is irrelevant. Assessment is important no matter whether it is used for improving teaching and learning or for evaluation (Brown & Gao, 2015). Research has also shown that for primary and secondary school teachers in New Zealand, there was a negative correlation between improvement and irrelevance, and a weak correlation between improvement and using assessment to evaluate students (Brown, 2004, 2011). New Zealand primary school teachers tended to associate improvement with school accountability and to moderately relate student accountability with irrelevance (Brown, 2004). In short, the aforementioned studies explored both the strength of agreement for the main conceptions of assessment held by teacher participants and the interrelation between them, which provided insights into teachers' conceptions of assessment. The current study also investigated these two issues related to Chinese EFL teachers' conceptions of assessment.

2.2. Chinese teachers' conceptions of assessment

The TcoA inventory has been applied to gauge Chinese teachers' conceptions of assessment. Since the four-factor framework could not capture the various conceptions held by Chinese teachers, Brown et al. (2011) created a TcoA inventory for Chinese contexts (C-TcoA) based on data collected from 1,014 primary and secondary school teachers in Hong Kong and 898 primary and secondary school teachers in Guangzhou. Three major interrelated factors have been identified based on teacher responses to a 6-point positively packed agreement rating scale (i.e., two negative and four positive rating points for each survey item to elicit variance in response to socially accepted statements, including *strongly disagree*, *mostly disagree*, *slightly agree*, *moderately agree*, *mostly agree*, and *strongly agree*). These three major factors include improvement, accountability, and irrelevance.

The improvement factor encompasses three sub-factors, that is, assessment is for student development, assessment is for helping students learn, and assessment results should be accurate. The accountability factor also consists of three sub-factors, that is, taking into account measurement errors in assessment use, using assessment to control teachers and evaluate schools, and utilizing examination as assessment. The irrelevance factor refers to the negative aspects of assessment.

Brown and Gao (2015) proposed a model of Chinese conceptions of assessment based on collaborative research between them and graduate student theses written under the supervision of Gao. The model includes six major conceptions, ranging from a more external management and control perspective to a more individualistic developmental view of assessment, in addition to a more negative view of assessment. These conceptions include management and inspection (i.e., using assessment to inspect and control schools, teachers, and students for better teaching and achievement); institutional targets (i.e., using assessment to check if students have achieved pre-set learning standards as instantiated in public examinations); facilitation and diagnosis (i.e., using assessment to provide valid information for the diagnosis and facilitation of teaching effectiveness); ability development (i.e., using assessment to increase students' motivation and learning abilities); personal quality (i.e., using assessment to enhance the overall quality of students); and negativity (i.e., assessment exerts a negative influence on teaching and learning).

Research on the C-TcoA has shown that Chinese teachers agreed most with the conception that assessment is needed for improvement (Brown et al., 2011; Chen & Brown, 2015). In Chen and Brown's (2015) study involving 1,500 Chinese teachers from primary, middle, and high schools, after "assessment as teacher improvement," "assessment is for student development" was the most endorsed view. A strong positive association was identified between assessment as improvement and assessment as accountability (Brown et al., 2011), indicating that teachers considered that examining students facilitated their learning. In Brown et al.'s (2011) study, a positive correlation was also found between assessment for accountability and irrelevance. In Chen and Brown's (2015) study, a moderately strong connection was found between school accountability and student development.

Despite the research on Chinese teachers' conceptions of assessment mentioned earlier (Brown et al., 2011; Chen & Brown, 2015), there is limited research on Chinese EFL (English as a foreign language) teachers' conceptions of assessment in the context of nationally mandated formative assessment innovation. Using the C-TcoA and assessment practices inventory (Zhang & Burry-Stock, 2003), Gan et al. (2018) probed into 107 Chinese secondary EFL teachers' conceptions and practices of assessment. Four main conceptions of assessment were identified, including "help learning," "student development," "teacher/student accountability," and

“examination and school accountability.” Like the teachers in other studies (Brown, 2011; Brown & Michaelides, 2011), the Chinese EFL teachers agreed most with the view that assessment helps students improve their learning. The second most endorsed view was “assessment as examination and school accountability.” A moderately strong correlation was identified between the “help learning” factor and the “student development” factor, and between the “teacher/student accountability” factor and the “examination and school accountability” factor. The “teacher/student accountability” factor was found to be weakly correlated to the “help learning” factor and the “student development” factor, respectively. A weak correlation was also identified between the “examination and school accountability” factor and the “student development” factor, while a medium level of correlation was found between the “examination and school accountability” factor and the “help learning” factor.

Gan et al.'s (2018) research also examined Chinese secondary EFL teachers' assessment practices. The teachers reported using different assessment practices frequently, including aligning teaching and assessment (e.g., matching assessment with instruction), using assessments for improvement (e.g., using assessment results when planning teaching), using traditional assessments (e.g., using multiple choice questions to assess students), sharing assessment criteria (e.g., communicating assessment criteria to students in advance), and providing oral feedback. However, the teachers seemed to only occasionally use student-centered assessments, such as self or peer assessment, a phenomenon also identified in other EFL contexts (e.g., Bui & Kong, 2019). The most frequently adopted assessment practice, aligning teaching and assessment, was associated with both the “help learning” factor and the “student development” factor, but not the “teacher/student accountability” factor, indicating that the teacher participants somehow implemented assessment-for-learning principles. Student-centered assessments were the only type of assessment that had no systematic relationship with the four main conceptions of assessment identified in Gan et al.'s (2018) study.

2.3. Factors affecting Chinese teachers' conceptions of assessment

Previous research utilizing C-TcoA has explained the teacher participants' conceptions of assessment through the influence of sociocultural and policy contexts. Chinese sociocultural values attach great importance to performance in public examinations, which informs decision-making regarding the selection of students for opportunities for better education (He et al., 2011). Public examination results are used to evaluate not only students, but also teachers and schools (Brown et al., 2011). At the same time, a person's academic achievement is also

associated with beliefs about personal worth and virtue (China Civilization Centre, 2007). Therefore, helping students achieve higher scores in public examinations not only contributes to their knowledge and performance, but also makes them better people (Brown et al., 2011). At the policy level, the current curriculum reform in China emphasizes an assessment reform, advocating the use of formative assessment in English language education to promote students' holistic development (Chinese Ministry of Education, 2017). According to Brown and Gao (2015), the assessment context seems to pull teachers towards different ends, that is, summative assessment emphasizing performance, and formative assessment emphasizing learning improvement.

Research has also shown that teacher characteristics (i.e., sex and teaching experience) may influence Chinese teachers' conceptions of assessment. For example, probably because more males assume the role of school leaders in Chinese schools (Brown & Gao, 2015), male teachers agreed more strongly with the management and inspection conception and the institutional targets conception (South China Normal University Team, 2010). Highly experienced teachers were found to agree more strongly with the management and inspection conception and the institutional targets conception, and to agree less with the personal quality conception and the facilitation and diagnosis conception (Brown & Gao, 2015).

Work environments constitute another source of influence. Teachers in senior secondary schools, who face the greatest pressure to prepare students to perform well in public examinations, agreed most with the irrelevance, management and inspection, as well as institutional targets conceptions, but agreed least with the personal quality conception (Wang, 2010). Teachers in the final year of senior secondary school agreed most with personal quality conception and those in higher ranking/banding schools agreed more with personal quality conception as well (Shang, 2007).

As can be seen from the literature review, research employing the TcoA has mainly adopted a quantitative approach to investigating conceptions of assessment held by teachers in different regions and countries (e.g., Brown, 2004, 2011; Brown & Michaelides, 2011; Chen & Brown, 2015; Gan et al., 2018), with the results being explained by sociocultural and policy contexts. Quantitative studies on factors affecting Chinese teachers' conceptions of assessment have also focused on particular categories of factors such as teacher characteristics and work environments (Shang, 2007; South China Normal University Team, 2010; Wang, 2010). The aforementioned research has contributed greatly to the understanding of teachers' views of assessment and factors affecting them. However, quantitative research can only reveal a general picture of teachers' conceptions of assessment without providing an in-depth understanding of the interaction among global and local factors in shaping individual teachers' views

and related practices of assessment. From an ecological perspective, teachers' assessment views and practices are influenced by three distinct but interacting levels of contextual factors, including macro-level factors (e.g., national and cultural influences), meso-level factors (e.g., school factors and expectations of parents and the immediate community), and micro-level factors (e.g., factors related to the classroom, students, and teachers), among which meso-level factors deserve more attention (Fulmer et al., 2015). To understand teachers' conceptions and practices of assessment in detail and in context, it seems that qualitative data should be utilized as well. This study utilized both quantitative and qualitative data for a more refined and contextualized understanding of Chinese teachers' conceptions of assessment in the context of the recent English language curriculum reform, which emphasizes formative assessment initiatives. If teachers do not endorse the view that assessment can be used to promote teaching and learning, as advocated in the education reform, then the proposed new form of assessment is unlikely to be successful. Sustainable assessment training programs are also needed to keep in-service teachers informed of assessment principles (Xu & Brown, 2017). However, attempting to change teachers' behaviors only (e.g., increasing formative assessment practices) without taking into consideration their existing beliefs is likely to fail (Brown & Gao, 2015). It is thus crucial to understand how Chinese EFL teachers conceive of assessment and factors affecting their conceptions both for the success of policy initiatives and teachers' professional development. Inspired by the research gaps identified in the literature review, this paper seeks to answer the following research questions:

- RQ1. What were the overall conceptions of assessment among the Chinese EFL teachers in the study, and what, if any, relations emerged among those conceptions?
- RQ2. What was the impact of teaching experience and school banding on the teacher participants' conceptions of assessment?
- RQ3. What were the individual teacher participants' conceptions and practices of assessment and what were the factors affecting them?

3. Methods

3.1. Research design

This study adopted a mixed-methods approach that involved both quantitative and qualitative data. An explanatory sequential mixed methods design (Creswell, 2014) was utilized. Quantitative data were collected first, followed by a qualitative phase of the study. The quantitative results informed the selection

of participants in the qualitative phase, with the qualitative data expecting to provide more depth and insights into the quantitative results of the study. To answer RQ1, the 31-item Chinese teachers' conceptions of assessment (C-TcoA) questionnaire was used to collect quantitative data to obtain a general picture of the teacher participants' views of assessment. As previous research identified the interrelationship among Chinese teachers' different conceptions of assessment (Brown et al., 2011; Gan et al., 2018), this study also aimed to examine whether the teacher participants' various views of assessment were potentially interrelated. To answer RQ2, the same set of quantitative data were utilized to ascertain the potential influence of teaching experience and school banding on the participants' conceptions of assessment, given that research has identified the influence of teacher characteristics (i.e., sex and teaching experience) and work environment (e.g., school banding) (Brown & Gao, 2015; Shang, 2007; Wang, 2010). We thus focused particularly on the two variables of teaching experience and school banding to identify their potential influence. Due to the very small number of male teachers in the study (i.e., 7 out of 66), the influence of sex on teacher conceptions of assessment was not investigated. Although the answer to RQ2 can shed light on the potential influence of micro-level factors (i.e., teaching experience as one teacher factor) and of meso-level factors (i.e., school banding as one school factor) on teachers' conceptions of assessment, in-depth qualitative data were needed to add to the quantitative data by exemplifying the potential interaction among macro-level, meso-level, and micro-level factors. Therefore, based on the findings of the first two research questions (i.e., the influence of school banding on the teacher participants' conceptions of using assessment to promote learning—see the section on results), two teachers from schools with different bandings were selected. Case studies of these teachers were conducted for RQ3 to understand their conceptions and practices of assessment in context and the different layers of shaping influences on them. In short, the mixed-methods approach allowed the investigation of a general tendency among a particular group of teachers and a contextualized understanding of individual teachers' assessment conceptions and practices.

3.2. Participants

For the quantitative part of the study, a purposive sample of 66 Chinese EFL teachers from six senior secondary schools in a city in Eastern China participated in the C-TcoA survey. These six schools were purposively selected based on two criteria. First, the schools represented different school bandings, including municipal-level key schools, district-level key schools, and general high schools. Secondary schools in China are categorized into those that enjoy higher banding or

reputation (i.e., key schools) and those that are not as reputable (i.e., non-key schools or general high schools) (Yu et al., 2016). Among the key schools, there is also a distinction between municipal-level key schools and district-level key schools, with the former being more prestigious than the latter. Second, the schools were known to the researchers. In this study, schools known to the researchers tended to be more supportive of the research project compared with those schools to be recruited from random sampling. Random sampling may be a relatively ineffective sampling strategy in Chinese school contexts (Brown et al., 2011). Table 1 shows the background information of the teacher participants.

Table 1 Background information of the teacher participants

Participants' background	Number (%)
Sex	
Female	59 (89.4%)
Male	7 (10.6%)
Educational background	
Bachelor's degree	41 (62.1%)
Master's degree	22 (33.3%)
Not given	3 (4.6%)
Teaching experience	
1-4 years	15 (22.7%)
5-18 years	18 (27.3%)
19-23 years	13 (19.7%)
Over 24 years	13 (19.7%)
Not given	7 (10.6%)
School banding	
General high schools	21 (31.8%)
District-level key schools	19 (28.8%)
Municipal-level key schools	25 (37.9%)
Not given	1 (1.5%)

The qualitative part of the study involved case studies of two purposefully selected teacher participants. A strength of case study is its capacity to provide an in-depth and contextualized understanding of contemporary real-life phenomena (Creswell, 2013). The teachers were chosen based on the following criteria: (1) they worked in schools with different bandings; (2) they were enthusiastic about and supportive of the research. As the quantitative analysis revealed that school banding (i.e., municipal-level key school vs. district-level key school) exerted an influence on teachers' conception of using assessment to promote learning (see the section on results), school banding was used as one of the criteria for case selection. Teacher A, a female teacher with 29 years of teaching experience, came from a municipal-level key school. Teacher B, a female teacher with 15 years of teaching experience at the time of study, came from a district-level key school.

3.3. Data collection and analysis

The quantitative data were mainly collected through the 31-item Chinese teachers' conceptions of assessment (C-TcoA) questionnaire (Brown et al., 2011), which helped to gauge the EFL teacher participants' conceptions of assessment. The C-TcoA elicited teachers' self-ratings for the following conceptions of assessment: (1) assessment helps teaching and learning; (2) assessment promotes students' development; (3) assessments are accurate; (4) assessment involves examinations; (5) measurement errors should be taken into consideration in assessment use; (6) assessment is used to control teachers and evaluate schools; and (7) assessments are irrelevant.

Confirmatory factor analysis was employed to determine if the EFL teacher participants' responses fitted the factor model identified by Brown et al. (2011) ($\chi^2/df = 1.70$, RMSEA = 0.10, RMR = 0.11, CFI = 0.94). As RMSEA¹ and RMR were greater than .08 and .05 respectively, exploratory factor analysis (EFA) was utilized to develop an alternative model. Prior to performing EFA, the suitability of data for factor analysis was assessed. The Kaiser-Meyer-Olkin value was .68 and Bartlett's test of sphericity reached statistical significance (approximate $\chi^2 = 725.27$, $df = 231$, $p = .00$), supporting the factorability of the correlation matrix. Varimax rotation was used for EFA. After EFA, inter-factor correlations were calculated to explore the potential relationships among the factors. As the data were not normally distributed, the Kruskal-Wallis test was used to examine the influence of: (a) teaching experience (1 to 4 years $N = 15$, 5 to 18 years $N = 18$, 19-23 years $N = 13$, over 24 years $N = 13$) and (b) school banding (general high school $N = 21$, district-level key school $N = 19$, municipal-level key school $N = 25$). Bonferroni correction was applied given that we ran two Kruskal-Wallis tests. Therefore, the threshold for the p value was set at $0.05/2 = 0.025$.

For the qualitative part of the study, two semi-structured interviews were conducted with two purposefully selected teachers to obtain a contextualized understanding of their conceptions and practices of assessment. The interviews were conducted in Chinese, the teachers' native language, but they were allowed to switch between Chinese and English whenever necessary for the sake of a clear expression of meaning. Each interview was audio recorded and lasted for about 45 minutes.

To analyze the interview data, we employed a qualitative data analysis scheme including data reduction, data display, and conclusion drawing and verification (Miles

¹ We decided to follow the guidelines endorsed in Brown (2015). That is, RMSEA values less than 0.05 suggest a good model fit; RMSEA values less than 0.08 suggest adequate model fit; RMSEAs in the range of 0.08-0.1 suggest a mediocre fit; and models with RMSEA value ≥ 0.1 should be rejected. Therefore, the RMSEA value of 0.10 in this study suggests an unsatisfactory model fit. The full results of RMSEA with the 90% CI statistics will be provided upon reader request.

et al., 2014). The interview data were transcribed verbatim and checked for accuracy. Data reduction was performed by treating a paragraph as a unit of coding and focusing on information reflecting the interviewees' conceptions and practices of assessment and factors affecting them. We used Brown and Gao's (2015) model of Chinese teachers' conceptions of assessment (i.e., management and inspection, institutional targets, facilitation and diagnosis, ability development, personal quality, and negativity) to code information related to conceptions of assessment. For example, the code "institutional targets" was assigned to the following data: "In my school, we mainly use tests to measure students' performance. The final grade is based on the average of students' test results." Regarding the coding of assessment practices, we utilized the six types of classroom assessment practices adopted by Chinese EFL teachers (Gan et al., 2018) as an analytical framework, which included aligning teaching and assessment, using assessments for improvement, using traditional assessments, sharing assessment criteria, providing oral feedback, and student-centered assessments. For instance, the code "using traditional assessments" was assigned to the following data: "Tests are conducted weekly, monthly, mid- and final-term. After test-taking drills and my explanation of the answers to the test, there is not much time left." We also coded information regarding the factors affecting the participants' conceptions or practices of assessment. For example, the code "influence of college entrance examination" was assigned to the following data: "If the college entrance examination is still used and if the English test paper is still so difficult, it is quite impossible to change the current situation." During data analysis, we were also open to new codes as well. The relationships between different codes were examined to develop emerging themes, such as the influence of college entrance examination on the use of traditional assessments. Case narratives were also developed for the teachers. Cross-case comparisons were conducted, with similarities and differences between cases identified and analyzed using matrixes. Conclusions about the teacher participants' conceptions and practices of assessment as well as factors affecting them were drawn and verified through member-checking.

To ensure the reliability and trustworthiness of data analysis, the two authors independently coded all the qualitative data and the inter-coder reliability reached 85%. They then discussed to resolve disagreements in coding. After a second round of coding, the inter-coder reliability reached 92%. Member-check interviews were also conducted to elicit the teachers' opinions on our interpretations of interview data.

4. Results

4.1. Teachers' conceptions of assessment: A general picture

RQ1 addressed the Chinese EFL teachers' conceptions of assessment and the interrelationship, if any, among the assessment conceptions. The revised C-TcoA model contained five inter-correlated factors (Table 2). Factor 1 (i.e., help learning), comprising 3 items, showed that assessment helps students to learn. Factor 2 (i.e., student/teacher accountability), containing 4 items, showed that teachers and students should be held accountable for teaching and learning. Factor 3 (i.e., assessment as accurate for student development), containing 5 items, identified assessment for student development. Factor 4 (i.e., assessment as accurate for examination and teacher/school control), containing 6 items, showed that assessment is used to prepare students for examinations and to control teacher and schools. Factor 5 (i.e., irrelevance), comprising 4 items, showed that assessment is irrelevant. Two of the factors identified by Brown et al. (2011) (i.e., help learning and irrelevance) were confirmed in the study.

Table 2 C-TcoA factors, items, and factor loadings based on exploratory factor analysis

Scale and items	Factor loading
Help learning	
1. Assessment helps students improve their learning.	.89
2. Assessment determines if students meet qualification standards.	.88
3. Assessment information modifies ongoing teaching of students.	.86
Student/teacher accountability	
22. Assessment sets the schedule or timetable for classes.	.62
23. Assessment helps students gain good scores in examinations.	.82
24. Assessment selects students for future education or employment opportunities.	.80
25. Assessment results contribute to teachers' appraisals.	.71
Assessment as accurate for student development	
4. Assessment results are sufficiently accurate.	.51
9. Assessment helps students succeed in authentic/real-world experiences.	.74
10. Assessment is used to provoke students to be interested in learning.	.77
11. Assessment cultivates students' positive attitudes towards life.	.67
13. Assessment stimulates students to think.	.67
Assessment as accurate for examinations and teacher/school control	
8. Assessment results can be depended on.	.56
14. Assessment is assigning a grade or level to student work.	.67
19. Assessment teaches examination-taking techniques.	.68
26. Assessment helps students avoid failures on examinations.	.61
6. Assessment is used by school leaders to police what teachers do.	.68
30. Assessment is an accurate indicator of a school's quality.	.45
Irrelevance	
12. Assessment results are filed and ignored.	.61
15. Assessment is an imprecise process.	.71
18. Assessment interferes with teaching.	.68
27. Assessment forces teachers to teach in a way against their beliefs.	.75

Table 3 C-TcoA factor means, SDs, and Cronbach's α

Factors	Number of items	Scale example	Cronbach's α	<i>M</i>	<i>SD</i>
1. Help learning	3	Assessment helps students improve their learning.	.90	4.93	1.27
2. Student/teacher accountability	4	Assessment selects students for future education or employment opportunities.	.82	3.66	1.04
3. Assessment as accurate for student development	5	Assessment cultivates students' positive attitudes towards life.	.81	4.20	0.85
4. Assessment as accurate for examination and teacher/school control	6	Assessment teaches examination-taking techniques.	.76	3.80	0.87
5. Irrelevance	4	Assessment forces teachers to teach in a way against their beliefs.	.71	3.12	1.13

Table 3 shows the mean score for each factor. The teacher participants tended to agree most with the conception that assessment is used to help learning. There was moderate agreement with the idea that assessment is for student development on condition that it is accurate. The teacher participants also tended to moderately agree that as long as assessment is accurate, it may be used to prepare students for exams and to control teacher/school and that students and teachers should be held accountable for assessment. The teachers slightly agreed that assessment is irrelevant.

Table 4 The inter-correlation between EFL teachers' assessment conception factors

Teacher assessment conceptions	1	2	3	4	5
1. Help learning	1				
2. Student/teacher accountability	-.12	1			
3. Assessment as accurate for student development	.36**	.28*	1		
4. Assessment as accurate for examination and teacher/school control	-.02	.48**	.55**	1	
5. Irrelevance	-.083	.45**	.05	.27*	1

* $p < .05$, ** $p < .01$

As indicated by Table 4, there was high inter-factor correlation between the "assessment as accurate for examination and teacher/school control" factor and the "assessment as accurate for student development" factor ($r = .55$). There was medium correlation between the "assessment as accurate for examination and teacher/school control" factor and the "student/teacher accountability" factor ($r = .48$), between the student/teacher accountability factor and the irrelevance factor ($r = .45$), and between the "help learning" factor and the "assessment as accurate for student development" factor ($r = .36$).

RQ2 investigated the influence of teaching experience and school banding on the teacher participants' conceptions of assessment. Regarding the influence of teaching experience, no statistically significant differences have been found across the four groups of teachers with different years of teaching experience.

Concerning the influence of school banding, a Kruskal-Wallis test revealed a statistically significant difference in the "help learning" factor across teachers from three types of schools with different bandings (municipal-level key schools, $N = 25$; district-level key schools, $N = 19$; general high schools, $N = 21$), $\chi^2(2, N = 65) = 8.124, p = .017$. The teachers from municipal-level key schools and general high schools both recorded median values of 6. The teachers from district-level key schools recorded a median value of 4. Mann-Whitney U tests further revealed a significant difference between the teachers from municipal-level key schools ($Md = 6, N = 25$) and those from district-level key schools ($Md = 4, N = 19$), $U = 128.5, z = -2.70, p = .007, r = .41$. In other words, teachers from municipal-level key schools seemed to agree more strongly than those from district-level key schools that assessment is for enhancing student learning.

4.2. Teachers' conceptions and practices of assessment: Two cases

RO3 probed into two individual teachers' conceptions and practices of assessment and factors affecting them. Interviews with the two teachers revealed individual differences in assessment conceptions and practices despite similarities. The two teachers' conceptions and practices of assessment are reported first, followed by a summary of factors affecting them.

Both teacher participants acknowledged that assessment may serve multiple purposes, but each highlighted different priorities. For example, Teacher A stated: "In my school, we mainly use tests to measure students' performance. The final grade is based on the average of students' test results." This quote reflected the conception that assessment is used as a mechanism to evaluate students. She added: "Assessment is mainly about giving tests to students, especially Senior Three students. As our school is a high-banding school, our school leaders want students to achieve high scores in external examinations, and teachers are forced to teach to the test. We don't have time to think about better ways to teach and to assess." This quote indicates that the teacher conceived assessment not only as administering tests to prepare students for external examinations such as the college entrance examination, but also as a mechanism by the school and school leaders to constrain what teachers do to raise students' examination scores, as can be seen from the use of the phrase "forced to teach to the test."

Teacher A expressed a sense of exhaustion by comparing the past and current situation: "In the past I could still decide what to teach in my class and I enjoyed teaching quite a lot, but in recent years the college entrance examination for the English subject has become more and more difficult, and I start to feel exhausted and I just want to retire. The examination has constrained what we have to teach." It seemed that Teacher A became less motivated to teach because the college entrance examination constrained what she could teach in class.

Concerning the most frequently used assessment practices, Teacher A thought that it was difficult to rank the different types of assessment practices as identified in Gan et al. (2018) because she stated that tests were used the most frequently in her English class, while student-centered assessment such as peer- or self-assessment was seldom used. She mentioned: "Tests are conducted weekly, monthly, mid- and final- term. After test-taking drills and my explanation of the answers to the test, there is not much time left." Although she was aware that peer- and self- assessment was promoted in the new senior secondary English language curriculum, she talked about the difficulty in implementing change: "If the college entrance examination is still used and if the English test paper is still so difficult, it is quite impossible to change the current situation." The quote indicated that from Teacher A's perspective the current examination system creates limited space for using formative assessment practices such as peer- or self- assessment.

In short, Teacher A regarded assessment as giving students, especially Senior Three students, tests to measure their performance and preparing them for the college entrance examination to achieve high scores and to fulfill school leaders' expectations. Her case suggested the influence of macro-level factor (i.e., the college entrance examination), meso-level factor (i.e., a high banding school with high expectations from school leaders), and micro-level factor (i.e., Senior Three students in a high banding school). Notably, although not explicitly mentioned by Teacher A, the students in her school were high achieving students compared with those from district-level key schools and general high schools (a point mentioned by Teacher B). They were thus expected to perform excellently in the college entrance examination.

Different from Teacher A, Teacher B talked about the formative assessment initiatives in the English education reform and highlighted the use of assessment for promoting learning and student development. To her, assessment meant the kind of classroom tasks students do and receive feedback on. She stated: "We create tasks for students to do in class, such as a group task for students to discuss themes in a piece of reading. I may provide feedback on different dimensions of the task such as verbal delivery, correctness of ideas, task fulfillment, and so on. I talk about the strengths and weaknesses, but more feedback is usually given to the weak group." Teacher B added: "We also have a combination of teacher-, self- and peer- assessment. For example, we may ask one group of students to peer assess another group. Although most of the time students only give marks, the more capable ones can provide comments too." These quotes suggested that the teacher conceived of the purpose of assessment as eliciting evidence that is subject to different sources of feedback, that is, the formative dimension of assessment.

Teacher B also commented on the affective aspect of teacher feedback: "Positive and accurate feedback can stimulate our students' interest in learning, which is an essential student quality. Encouragement and guidance help students make progress not only in their academic study, but also in their life." This suggested that the teacher considered assessment to promote students' development through positive and to-the-point teacher feedback. She explained: "The students need a teacher who can guide not only their academic study, but also their views of the world and life."

Regarding the most frequently used assessment practices, teacher oral feedback and student-centered assessment (e.g., peer- and self-assessment) were regarded as the top two most frequently used practices in Teacher B's English classes. Using traditional assessment methods such as tests was ranked as the least used type. Teacher B explained: "School leaders in reputable schools may have high expectations on their teachers regarding the admission of students into prestigious universities, and this may give teachers great pressure to prepare students for external examinations. They are in a cycle of giving students tests and then explaining test answers. In our school, the most important task is to raise our students' interest in English and foster positive learning attitudes, particularly in the first two years of senior high school. This is because our students are not as good as those in reputable schools." Teacher B explained that although she came from a district-level key school, the students in her school were similar to those from general high schools in terms of academic performance.

Overall, Teacher B regarded assessment as a means of promoting student learning and development. In particular, she underscored the importance of providing feedback on students' task performance and using it to encourage and guide her students, particularly for Senior One and Two students. Despite the fact that she worked in a district-level key school, her students resembled those from general high schools academically. Therefore, her top priority seemed to be the use of feedback to motivate and promote students' learning during their senior one and two study, with the awareness that her practices were consistent with the formative view of assessment as advocated in the English curriculum reform. Teacher B's case reflected the influence of meso-level (i.e., school banding), micro-level (i.e., average performing students studying in senior one and two in a less prestigious school), and macro-level factors (i.e., the formative assessment initiatives in the English education reform) on her views of assessment, although the other macro-level factor (i.e., the college entrance examination) remained the same for her school.

Table 5 summarizes the two teacher participants' conceptions of assessment with reference to Brown and Gao's (2015) framework.

Table 5 A comparison between the two teachers' conceptions of assessment

Brown and Gao's (2015) framework	Teacher A	Teacher B
Management and inspection	✓	
Institutional targets	✓	
Facilitation and diagnosis		✓
Ability development		✓
Personal quality		✓
Negativity		

5. Discussion

This study has sought to answer three research questions related to Chinese secondary EFL teachers' conceptions of assessment. Regarding RQ1, the study has identified five major conceptions of assessment among the Chinese EFL teachers based on the Chinese teachers' conceptions of assessment inventory (Brown et al., 2011). The "help learning" factor referred to using assessment to improve learning and teaching and determine if students meet qualification standards. The "assessment as accurate for student development" factor indicated that as long as assessment results are sufficiently accurate, assessment helps students succeed in real-life experiences, stimulates their thinking and interest in learning, and cultivates their positive attitudes toward life. The "assessment as accurate for examination and teacher/school control" factor suggested that as long as assessment results are reliable, it can be used to prepare students for examinations, control what teachers do, and indicate a school's quality. The "student/teacher accountability" factor suggested that assessment selects students for future education or employment opportunities and assessment results contribute to teachers' appraisals. The "irrelevance" factor meant that assessment is an imprecise process, interferes with teaching, forces teachers to teach in a way against their beliefs, and assessment results are ignored. The "help learning" factor and the "student/teacher accountability" factor were consistent with Gan et al.'s (2018) research on Chinese EFL teachers. The "assessment as accurate for student development" factor and the "assessment as accurate for examination and teacher/school control" factor were different from their study. This group of teacher participants bundled the notion that assessment is accurate and reliable with both "student development and examination" and "teacher/school control." It seemed that to the teacher participants, judgments about student development as well as examination preparation and the control of teacher/school depend on whether assessment is accurate and reliable. The "irrelevance" factor identified in the study was not found in Gan et al.'s (2018) study. In the study, the most endorsed conception was that assessment is used to help learning. In this sense, this group of teachers held similar

views to those in previous research investigating Chinese secondary EFL teachers (Gan et al., 2018), New Zealand secondary school teachers (Brown, 2011), and Cypriot teachers (Brown & Michaelides, 2011). However, the teacher participants were different from the Chinese teachers in Brown et al.'s (2011) research where the same inventory was used.

There was strong inter-correlation between the "assessment as accurate for examinations and teacher/school control" factor and the "assessment as accurate for student development" factor ($r = .55$). In other words, as long as assessment is accurate, using assessment to prepare students for examinations and to control teachers/schools may also facilitate students' development. Such an association can probably be explained by the Chinese idea that excellent assessment results reflect a more valuable person (Brown et al., 2011). In the Chinese context, one who achieves good scores in examinations is regarded as a good person because examination results indicate the quality and worth of the individual (China Civilization Centre, 2007).

There was medium correlation between the "assessment as accurate for examinations and teacher/school control" factor and the "student/teacher accountability" factor ($r = .48$) in the teachers' conceptions of assessment. This indicated that those teachers who regarded assessment as a mechanism to evaluate teachers and students also considered it to be a way to prepare students for examinations and to control teachers and schools on condition that it is accurate. The Chinese society attaches great importance to public examination results because they are utilized to select students and evaluate teachers and schools (Brown et al., 2011). Therefore, schools, teachers, and learners face great pressure to ensure that students perform well in external high-stakes examinations. More often than not, drilling test-taking skills is employed for that purpose. For example, as mentioned by Teacher A, her lesson was dominated by the practice of test-taking skills because she was under school pressure to produce high-achievers in the English test of the college entrance examination.

There was also medium correlation between the "student/teacher accountability" factor and the "irrelevance" factor ($r = .45$). This suggested that when it is connected to student/teacher accountability, assessment is likely to be irrelevant. While this finding was not reported in Gan et al.'s (2018) study, it was somewhat similar to the finding in Brown's (2004) research on New Zealand primary school teachers. It should be noted that only student accountability was moderately related to irrelevance in Brown's (2004) study, while in this study both teacher and student accountability was associated with irrelevance. The teacher participants questioned the validity of assessment as teacher and student accountability probably because they were less convinced that public examination results alone can account for either students' quality of learning or

teachers' quality of teaching. For example, as mentioned by teacher B: "examination results cannot fully reflect teaching or learning quality."

A medium-strength correlation was also found between the "help learning" factor and the "assessment as accurate for student development" factor ($r = .36$). The finding indicated that assessment, perceived to contribute to learning, is also considered to facilitate student development if it is accurate. Teacher beliefs may be subject to the influence of historical, social, cultural, and policy contexts (Brown et al., 2019). Chinese teachers adhere to the cultural value that being a teacher involves educating students in not only the academic dimension, but also attitudinal and behavioral dimensions. This cultural value is reflected by the meaning of "cultivating" in Chinese (Gao & Watkins, 2001) and the Chinese expression "Jiao Shu Yu Ren," which means imparting knowledge and educating students to be good people in the society. Just as teacher B pointed out: "The students need a teacher who can guide not only their academic study, but also their views of the world and life." The current educational policy in China emphasizing students' holistic development, including linguistic development, cultural awareness, moral development, and thinking and learning skills (Chinese Ministry of Education, 2017), may be another reason for the connection between the "help learning" conception and the "assessment as accurate for student development" conception.

Regarding RQ2, this study has identified the influence of school banding on teachers' conception of assessment as helping with learning. Teachers from municipal-level key schools agreed more strongly with the idea that assessment is to promote learning compared with those from district-level key schools. While previous research showed that Chinese teachers in high-status/banding secondary schools agreed more with personal quality factors (Shang, 2007), this study further revealed that work environment such as school banding may influence Chinese teachers' conceptions of assessment related to using assessment to enhance learning. Such an influence indicated the need to take into consideration the meso-level factor of school environment (i.e., school banding) in relation to the implementation of formative assessment initiatives and teacher assessment training.

To sum up, the quantitative data revealed a general picture of the Chinese secondary EFL teachers' conceptions of assessment. Macro-level factors (sociocultural and policy contexts) were used to explain the connection between their different conceptions of assessment. The quantitative data also demonstrated the impact of one meso-level factor (i.e., school banding) on the teachers' conceptions of assessment.

Regarding RQ3, the qualitative data further identified the differences in two individual teachers' conceptions and practices of assessment. It seemed that the conceptions of Teachers A and B represented opposite points in the continuum

describing Chinese teachers' thinking of assessment (Brown & Gao, 2015). That is, Teacher A's views indicated the management and inspection (e.g., using assessment to control teachers so as to urge better achievement) and institutional target (e.g., using assessment to measure students' performance and to prepare them for examinations) parts of the continuum. Teacher B's views, on the other hand, suggested the facilitation and diagnosis (e.g., providing oral feedback on students' performance), ability development (e.g., using positive teacher feedback to motivate students), and personal equality (e.g., using teacher feedback to guide students' views of the world and life) parts of the continuum. In general, Teacher A's and Teacher B's conceptions of assessment reflected the summative (e.g., summative examination and judgment of learner outcomes) and formative (e.g., feedback provision, improved learning and learning motivation) dimensions of assessment, respectively. In accordance with the different conceptions of assessment, the two teachers prioritized either summative or formative assessment practices in their English classes.

The aforementioned differences can largely be attributed to the role of a meso-level factor (i.e., school factor) and related to it, a micro-level factor (i.e., student factors) in mediating the influence of macro-level factors (sociocultural and policy contexts) to shape teachers' different conceptions of assessment towards either the summative or formative end of the continuum. The assessment context in China may push teachers towards two different ends of the assessment continuum (i.e., the summative or formative ends) (Brown & Gao, 2015). As high-stakes test may stimulate intensive test preparation in the classroom (Qi, 2004), Teacher A's assessment conceptions and practices can be said to be derived from the washback effect of the college entrance examination. However, in the study it was the interplay of various contextual factors that contributed to her conceptions and practices of assessment. Teacher A's school context (i.e., reputable school, school leaders' high expectations of teachers and students) and the high achieving Senior Three students studying in it reinforced summative views of assessment predominant in sociocultural values (i.e., the importance of the college entrance examination). Teacher B's school context (i.e., a school with a lesser reputation, less pressure from leaders) and its average-performing Senior One and Two students seemed to be more conducive to fostering her learning-focused views of assessment as advocated in the English curriculum reform document (Chinese Ministry of Education, 2017), despite the importance of the college entrance examination.

According to Fulmer et al. (2015), meso level factors and their connection with macro- or micro-level factors are worth attention in research on teachers' assessment conceptions, knowledge, or practices. As demonstrated by the quantitative part of the study, a meso-level factor (i.e., school banding) exerted

an influence on Chinese secondary EFL school teachers' conceptions of assessment. The qualitative part of the study further identified the role of a meso-level factor (e.g., school banding) and a micro-level factor (e.g., the kind of students in schools with different bandings) in mediating macro-level factors (e.g., the college entrance examination). The qualitative findings showed the interaction among the meso-level, micro-level and macro-level factors in explaining individual Chinese secondary EFL teachers' conceptions of assessment. Notably, while the quantitative data showed that teachers in municipal-level key schools agreed more than those in district-level key schools that assessment is for promoting learning, the qualitative data showed a different pattern in the two individual teachers' conceptions. This contrast between the quantitative and qualitative findings was probably due to the fact that the former reflected the general tendency of teachers as groups (i.e., groups of teachers from municipal-level or district-level key schools), while the latter revealed the conceptions of assessment held by teachers as individuals because of the interplay among macro-, meso-, and micro-level factors. Such a contrast highlighted the importance of using qualitative data to add to quantitative data for an in-depth understanding of teachers' conceptions of assessment, which is subject to various layers of contextual factors.

Concerning the implications of the study, the Chinese secondary EFL teachers as a group associated examination and teacher/school control with student development, which makes it less likely for the teachers to adopt formative assessment initiatives that aim to foster students' holistic development as mandated by the English curriculum reform (Chinese Ministry of Education, 2017). As pointed out by Brown et al. (2011), if a relevant accountability authority places much less emphasis on employing high-stakes examinations to evaluate students, then changes in teacher beliefs and practices are much more likely. This point has also been echoed by teacher A. In China (e.g., Zhejiang and Jiangsu Provinces) there has been a recent attempt to reform the college entrance examination by including more criteria for university admission (e.g., personal growth portfolios) in addition to examination scores (Gan et al., 2018). However, at the current stage, public examinations still dominate the educational context, and there may be difficulties for the Chinese EFL teachers in the study to embrace formative assessment emphasizing students' holistic development.

Although the teachers in the municipal-level key schools as a group tended to endorse the view that assessment is used to enhance learning, Teacher A's case indicated that in the same high banding school there may be individual teachers like her who believed less in the idea of using assessment for learning due to the interactional impact of meso-level factors (e.g., school banding), micro-level factors (e.g., student factors) and macro-level factors (e.g., the college entrance examination). Her case suggested that a situated approach

should be adopted to introduce changes into the assessment beliefs and practices of teachers such as her. Such an approach is a complex endeavor which involves the consideration of the three layers of factors as mentioned earlier. For example, although limited changes can be made to macro-level factors (e.g., the college entrance examination) currently, meso-level factors can be manipulated to influence the assessment conceptions and practices of teachers like Teacher A. As the opportunities for reflective practices and participation in learning communities represent two main ways of teacher learning to enhance teachers' assessment literacy (Xu & Brown, 2016), school leaders may establish a community of practice (Wenger-Trayner & Wenger-Trayner, 2015) comprising leaders and teachers who share the same visions regarding the learning purposes of assessment. Such a community may then promote a formative view of assessment to teachers such as Teacher A and gradually involve them in participating reflectively in the community of practice. Notably, in an attempt to create such a facilitative school environment, school leaders themselves need to first reflect on their views of assessment and obtain more knowledge about formative assessment. Since the aforementioned meso-level factor will also interact with micro-level factors (e.g., the summative views of assessment already held by Teacher A), it is important to promote a form of formative assessment that teachers may find contextually appropriate (e.g., formative use of summative assessment in Teacher A's case) to influence their conceptions of assessment towards the formative end of the continuum.

Compared with their counterparts in municipal-level key schools, the teachers in the district-level key schools overall endorsed less the view of using assessment for learning purposes. However, Teacher B's case suggested that a formative view of assessment can be fostered due to an interplay of macro-, meso- and micro-level factors. In schools such as the one where Teacher B worked, a situated approach to shaping teachers' assessment conceptions and practices can also be adopted. Despite the fact that few changes can be made to the macro-level factors (e.g., the college entrance examination), at the school level (i.e., meso-level), a community of practice involving teachers such as Teacher B as key members can be built and opportunities should be given to these teachers to share with their colleagues the formative views and practices of assessment, with the aim of involving the reflective participation of more teachers in the community. To improve the effectiveness of such sharing activities, it is important to pay attention to not only the key members' assessment conceptions, but also their assessment knowledge (i.e., micro-level factors). In this way, adequate suggestions on different types of contextually appropriate formative assessment can be provided to different kinds of teachers according to their micro-level factors (e.g., those teaching Senior One and Two versus those teaching Senior Three).

In this sense, teachers such as Teacher B need to further enhance their knowledge of formative assessment, despite the formative view of assessment and awareness of its cognitive and affective benefits. For example, Teacher B believed that formative assessment was reserved for average-performing students like those in her school who needed more teacher scaffolding and encouragement, and that high achieving students in Teacher A's school did not need it. Formative assessment is powerful in improving weak students' performance (Black & William, 1998), but it does not mean that it should only be reserved for average or weak students. In addition, Teacher B seemed to attach less importance to using assessment results to inform instruction, despite her use of teacher oral feedback and student-centered assessment practices. This lack of connection between assessment and instruction has also been identified in Lam's (2019) research on Hong Kong secondary English teachers. Teacher B's case showed that demonstrating formative conceptions of assessment does not necessarily mean that the teacher has sophisticated and sufficient knowledge of formative assessment. If teachers like her have to play a key role in sharing their formative conceptions and practices of assessment and encouraging colleagues to participate in the community of practice, it is necessary to ensure that they possess appropriate conceptions as well as knowledge of formative assessment.

6. Conclusion

This study has sought to explore Chinese secondary EFL teachers' conceptions of assessment and the shaping influences on it based on both quantitative and qualitative data. As a group, the teacher participants agreed most strongly with the view that assessment is used to promote learning. However, the strong association they made between the "assessment as accurate for examination and teacher/school control" factor and the "assessment as accurate for student development" factor suggested that the formative assessment initiatives focusing on students' holistic development as promoted in the English curriculum reform are less likely to be adopted by the teachers as a group at the current stage. The quantitative analysis also identified the influence of one meso-level factor (i.e., school banding) on the teachers' conception of assessment as helping with learning. Qualitative data further demonstrated how a meso-level factor (e.g., school factors such as school banding) and a micro-level factor (e.g., student factors) interacted with each other to mediate the macro-level factor (e.g., the college entrance examination) in shaping Teacher A's and Teacher B's conceptions of assessment, representing the summative and formative dimensions of assessment, respectively.

This study has demonstrated the importance of utilizing both quantitative and qualitative data to provide the general pattern and contextualized understanding of

Chinese secondary EFL teachers' conceptions of assessment. In particular, the qualitative data added to the quantitative data by demonstrating the situated nature of teacher conceptions of assessment, which are subject to the interaction of various contextual factors. Accordingly, a situated approach paying special attention to the interacting impact of meso-level (i.e., school factor) and micro-level factors (e.g., teacher and student factors) should be adopted to shape the teachers' views and knowledge of assessment and to facilitate the implementation of formative assessment as advocated in English curriculum reform in China. This study only involved a purposive sample of 66 teachers from six secondary schools in Eastern China, so its findings can only be generalized to similar contexts. Nevertheless, the investigation has shown the importance of considering the interplay of macro-, meso- and micro-level factors in exploring teachers' conceptions of assessment through a mixed-methods approach and proposed a situated approach to developing teachers' assessment literacy. Future research may involve a more representative sample with the use of both perception and classroom observation data to explore EFL teachers' conceptions of assessment. Research may also investigate effective ways to implement formative assessment at the classroom and school levels based on a situated approach.

References

- Berry, R., & Adamson, B. (Eds.). (2011). *Assessment reform in education: Policy and practice* (Vol. 14). Springer Science & Business Media.
- Black, P., & William, D. (1998, October). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-149. <https://doi.org/10.1177/003172171009200119>
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, *11*, 301-318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters*, *3*, 45-70. <https://doi.org/10.18296/am.0097>
- Brown, G. T., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education*, *2*(1), 993836. <https://doi.org/10.1080/2331186x.2014.993836>
- Brown, G. T., Gebriel, A., & Michaelides, M. P. (2019). Teachers' conceptions of assessment: a global phenomenon or a global localism. *Frontiers in Education*, *4*(16). <https://doi.org/10.3389/educ.2019.00016>
- Brown, G. T. L., Hui, S. K. F., Yu, F. W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, *50*, 307-320. <https://doi.org/10.1016/j.ijer.2011.10.003>
- Brown, G. T. L., & Michaelides, M. P. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education*, *26*(3), 319-337. <https://doi.org/10.1007/s10212-010-0052-3>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.
- Bui, G., & Kong, A. (2019). Metacognitive instruction for peer review interaction in L2 writing. *Journal of Writing Research*, *11*(2), 357-392. <https://doi.org/10.17239/jowr-2019.11.02.05>
- Chen, J., & Brown, G. T. (2016). Tensions between knowledge transmission and student-focused teaching approaches to assessment purposes: Helping students improve through transmission. *Teachers and Teaching*, *22* (3), 350-367. <https://doi.org/10.1080/13540602.2015.1058592>
- China Civilization Centre. (2007). *China: Five thousand years of history and civilization*. City University of Hong Kong Press.
- Chinese Ministry of Education. (2017). *English curriculum standards for senior high schools*. People's Education Press.

- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. Sage.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
- Fives, H., & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers’ beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Individual differences and cultural and contextual factors* (Vol. 2, pp. 471-499). American Psychological Association.
- Fulmer, G. W., Lee, I. C., & Tan, K. H. (2015). Multi-level model of contextual factors and teachers’ assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 475-494. <https://doi.org/10.1080/0969594x.2015.1017445>
- Gan, Z., Leong, S. S., Su, Y., & He, J. (2018). Understanding Chinese EFL teachers’ conceptions and practices of assessment: Implications for teacher assessment literacy development. *Australian Review of Applied Linguistics*, 41(1), 4-27. <https://doi.org/10.1075/aral.17077.gan>
- Gao, L., & Watkins, D. (2001). Identifying and assessing the conceptions of teaching of secondary school physics teachers in China. *British Journal of Educational Psychology*, 71(3), 443-469. <https://doi.org/10.1348/000709901158613>
- Hao, J., & Otani, M. (2016). English education in high schools in China: Its current status and problems. *Memoirs of the Faculty of Education of Shimane University (Educational Science)*, 50, 65-73.
- He, Y., Levin, B. B., & Li, Y. (2011). Comparing the content and sources of the pedagogical beliefs of Chinese and American pre-service teachers. *Journal of Education for Teaching*, 37, 155-171. <https://doi.org/10.1080/02607476.2011.558270>
- Kennedy, K. J., & Lee, J. (2008). *Changing schools in Asia: Schools for the knowledge society*. Routledge.
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78-89. <https://doi.org/10.1016/j.system.2019.01.006>
- Marton, F. (1981). Phenomenography – describing conceptions of the world around us. *Instructional Science*, 10 (2), 177-200.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Sage Publications.
- Pajares, M. F. (1992). Teachers’ beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62, 307-332. <https://doi.org/10.3102/00346543062003307>

- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 147-170). Lawrence Erlbaum.
- South China Normal University Team. (2010, July). *Teachers' conceptions of assessment: Developing models for teachers in China* [Paper presentation]. The International Test Commission Conference 2010, The Chinese University of Hong Kong, Shatin. <http://www.itc2010hk.com/>
- Shang, H. (2007). *Research on the middle school teachers' conceptions of learning assessment* (Unpublished master's thesis). South China Normal University, Guangzhou.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 133-146). Sage Publications.
- Teng, F., & Bui, G. (2020). Thai university students studying in China: Identity, imagined communities, and communities of practice. *Applied Linguistics Review*, 11(2), 341-368. <https://doi.org/10.1515/applirev-2017-0109>
- Wang, P. (2010). *Research on the Chinese teachers' conceptions and practice of assessment* (Unpublished doctoral dissertation). South China Normal University, Guangzhou (in Chinese).
- Wenger-Trayner, E., & Wenger-Trayner, B. (2015). *Introduction to communities of practice: A brief overview of the concept and its uses*. <https://wenger-trayner.com/introduction-to-communities-of-practice/>
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Xu, Y., & Brown, G. T. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133-158.
- Yu, C., Wei, F., Li, L., Morrissey, P., & Chen, N. (2016). *Social attitudes in contemporary China*. Routledge.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342. https://doi.org/10.1207/s15324818ame1604_4