



EFL Teachers' Online Assessment Practices During the COVID-19 Pandemic: Changes and Mediating Factors

Cong Zhang¹ · Xun Yan² · Junju Wang¹

Accepted: 2 May 2021 / Published online: 23 May 2021
© De La Salle University 2021

Introduction

The ability to develop and use assessments is a key construct of professional development for language teachers because they are frequently involved in summative and formative assessments in the school settings (Author, 2018). Previous research shows that language teachers' assessment practices are influenced by both contextual and experiential factors (Author, 2020; Crusan et al. 2016). While most studies focus on assessment practices in face-to-face classrooms, few studies have examined teachers' assessment practices online. The outbreak of COVID-19 pandemic has led to a shift in teaching and assessment from the face-to-face mode to the online mode. This shift can have a long-lasting impact on classroom-based assessment even after the pandemic. Thus, understanding teachers' online assessment practices and what factors influence their practices is important for not only assessment trainers, but also language educators and policy makers in TESOL and Applied Linguistics. Situated in China's mainland, this study investigates the online assessment practices during the pandemic of six English as a Foreign Language (EFL) teachers in a Chinese university. We qualitatively explored what changes they made to adjust to the needs of the new

assessment mode and what factors mediated their changes of practices.

Literature Review

Language Teachers' Assessment Practices: Impact of Contextual and Experiential Factors

In the field of language assessment, the ability to develop and use assessments has been termed language assessment literacy (LAL). Davies (2008) categorized three elements of LAL: skills, knowledge, and principles. The construct was further expanded to include sociocultural, sociopolitical and historical dimensions (Fulcher, 2012). Taylor (2013) hypothesized a set of distinct LAL competence profiles based on the assessment needs and expectations for different stakeholders, wherein for language teachers, LAL places more weights on language pedagogy and local assessment practice than on knowledge of theory and principles and concepts in assessment. In fostering LAL among language teachers, while earlier studies emphasized the role of formal training in building the assessment foundations for language teachers (Brown & Bailey, 2008; Pill & Harding, 2013; Xu & Brown, 2017), more recent studies have focused on the engagement with assessment practices as a means to gradual LAL development (Author, 2018; Kleinsasser, 2005; Mertler, 1999).

Language teachers' assessment practices can be influenced by both contextual and experiential variables. Author (2018) defined contextual factors as larger educational, social, political, historical, or other factors, and experiential factors as assessment background, training, and practice. Previous research has indicated that contextual factors can collectively form an assessment culture that influences teachers' assessment practices in a local context (e.g., Rea-

✉ Junju Wang
wangjunju@sdu.edu.cn

¹ School of Foreign Languages and Literature, Shandong University, Jinan, China

² Department of Linguistics, University of Illinois at Urbana-Champaign, Urbana, USA

Dickins, 2001; Vogt & Tsigari, 2014), in particular, the development and use of assessments in school settings. In addition, research on experiential factors suggests that (1) teachers are more likely to use assessment practices they are familiar with (Reynolds-Keefers, 2010; Rohl, 1999); and (2) in case of new assessment activities, methods, or tools, they can learn about assessment on the job, through which they develop assessment intuitions (Scarino, 2013; Vogt & Tsigari, 2014).

Impact of Policies on Language Teaching, Learning, and Assessment

The impact of policies on language teaching, learning, and assessment cannot be overstressed. Educational and language policy that has an impact on language assessment can vary in terms of scope, ranging from the introduction of a new form of testing at a local school, to a change in the teaching and assessment modalities to accord with changes to the standardized curriculum at the national level. As McNamara (2011) argues, language assessments often play a mediating role between educational policies and teaching and learning, the needs and philosophies of which are often at odds to each other; balancing between the two can be a challenging task for assessment developers, especially when language teachers have to do the assessing. As such, new policies not only change the functions of language assessments in school settings, but also redefine relationships among assessments, curriculums, and pedagogical practices (Harper et al. 2007; Jin et al. 2017; North, 2000).

The outbreak of the coronavirus disease in 2019, a.k.a. COVID-19 pandemic, has created profound impact on language education, which has traditionally relied on face-to-face instruction and assessment. Although educational policies have been made to accommodate the needs for alternative pedagogical and assessment methods, the quick spread of the disease leaves language teachers and testers little time to contemplate the alignment among language policies, teaching, and assessment. Thus, a great degree of variability in classroom-based assessment can be expected across language teachers during this challenging time. Examining such variability is important for second and foreign language education as it contributes to a better understanding of the impact of educational policy on language assessment practices for teachers and the range of alternative assessment practices in the classroom.

Motivated by the aforementioned gaps, this study employs a qualitative approach to closely examine the online assessment practices of EFL teachers in China during the COVID-19 pandemic. Using semi-structured retrospective interviews with six informants, this study addresses the following research questions:

- (1) What changes did Chinese EFL teachers make in online assessment during the COVID-19 pandemic?
- (2) What leads to the changes in their assessment practices?

Methodology

Research Context

This study took place in a large public university (University A) in Eastern China and investigated the online assessment practices during the pandemic of EFL teachers. Due to the outbreak of the pandemic, the Ministry of Education of China mandated that universities offer online courses during the semester of Spring 2020. The policy of turning all teaching online unavoidably influenced the way of assessment. In addition to replacing face-to-face instruction, University A implemented another policy that the percentage of formative assessment should be raised and that of summative assessment be decreased. However, when implementing these top-down policies, the government or the universities did not have enough time to plan; nor did they have prior experience to refer to. Thus, teachers had to make changes based on the local context and their own experiences and intuitions, and their changes of assessment practice were likely to display individual differences.

Participants

Participants were recruited through purposive and convenience sampling. Since we want to explore the online assessment practices of EFL teachers, we specifically sent an invitation email to the School of Foreign Languages and Literature at University A that one of the authors had connection with. Six teachers responded to this invitation and agreed to participate the study. The participants' (pseudonyms) profiles are shown in Table 1. Three participants were male and three were female. Except for one who held a master's degree and was completing his PhD, all the other participants held a PhD degree. In terms of ranking, two participants were lecturers, three were associate professors, and one was a professor. Except for Henry who was a tenure-track position lecturer (a.k.a. "Qingjiao" in Chinese), all the other participants held permanent positions in University A. Their age ranged between 34 and 49 years old, and their teaching experiences ranged from five to 24 years. The courses they taught included college English, writing, translation, speaking and comprehensive

¹ Some of the participants taught more than one course, and they picked one course to report their online assessment practices.

Table 1 Interviewees' profile

Name	Gender	Age	Degree	Title	Course Taught	Years of teaching
Steve	Male	40	MA	Lecturer	Writing	15
Henry	Male	34	PhD	Lecturer	Writing	5
Helen	Female	45	PhD	Associate Professor	Speaking	20
Cathy	Female	48	PhD	Associate Professor	College English	24
Mary	Female	49	PhD	Professor	Translation	24
Jim	Male	46	PhD	Associate Professor	Comprehensive English	21

English.¹ Among them, half of the participants claimed prior knowledge with language assessment at the time of data collection. Steve took courses in language testing during his graduate study and claimed to be interested in classroom-based assessment. Cathy had been a certified rater of a national English test for over 12 years. During the years, she had received some training from the center of the test. Mary took a language assessment course during her PhD study about 10 years ago. The diversity of the participants in gender, age, degree, rank, and teaching and assessment experience added the richness of the data for this study.

Data Collection and Analysis

Data was collected through semi-structured retrospective interviews. An interview protocol (see Appendix) was used to guide the interviews and contained four parts. The first part inquired about the demographic information of the participants, including their age, degree, rank, courses taught during Spring 2020 and years of teaching at the university. The second part focused on the participants' online assessment practices. The third part asked for their change of assessment practices from offline (before the pandemic) to online (during the pandemic) and the reasons behind the changes. The last part was about their assessment knowledge and theory. When conducting the interviews, the interviewer followed the protocol, and asked follow-up questions based on the interviewees' responses. At the request of the participants, all the interviews were conducted in Chinese, the first language of the participants. Each interview was conducted face-to-face, lasting between 45 and 70 min. All interviews were audio-recorded and transcribed verbatim, yielding a total of 114,458 Chinese characters. The interview transcripts were shown to the interviewees to ensure the authenticity and reliability of the interview data.

When doing data analysis, the original Chinese interview transcripts were used to ensure that no important information was missing due to translation. The transcripts were translated only when they were used to present results. The interview transcripts were iteratively coded

and analyzed in NVivo 11 (QSR 2012). Thematic analysis was used to code the data and Bryman's (2015) data analysis steps were followed. First, one author read all the data to identify relevant opinions, group similar codes into a larger overarching category, and then look for patterns among or between categories. Once the initial coding scheme was formed, the researchers applied the coding scheme to all the data. The coding scheme (see Table 2) went through several rounds of changes and modification in the process of iterative coding. We also double-coded the data to ensure inter-coder reliability. During this process, one difficulty we faced was to identify the factors that influenced the changes in the participants' assessment practices. To do this, we re-read the transcripts for multiple times and independently identified all possible factors associated with the assessment practices of each participant. The inter-code agreement was 0.796 for "Assessment context", 0.764 for "Experiential factors", and 0.872 for "Assessment changes". Then, discrepancies in our codes were discussed, especially the first two categories, until we reached consensus on all the codes.

Results

Due to the outbreak of COVID-19 pandemic, the Ministry of Education of China mandated that universities offer online courses during the semester of Spring 2020. Along with this policy, University A required teachers to raise the percentage of formative assessment and reduce the percentage of summative assessment. As for the specific way of assessment, University A did not offer detailed methods and left the freedom to teachers. Under this policy, teachers made various changes mediated by different factors, as shown in the reports below.

Jim and Helen

Jim and Helen did not make many changes to their assessment practices during teaching online. Jim only raised the ratio of formative assessment from 20 to 40%, as mandated by the university's policy. Although Jim agreed

Table 2 Coding scheme for semi-structured interviews with the teachers

Category	Subcategory	Topics	Codes
Assessment context	National context	COVID-19 pandemic	1-a-i
		National policy	1-a-ii
		National resources	1-a-iii
	Institutional context	Institutional policy	1-b-i
		Institutional support	1-b-ii
	Local context	Class size	1-c-i
		Internet connection	1-c-ii
		Technology	1-c-iii
	Experiential factors	Motivation	
Attitudes toward new assessment methods			2-b
Ability to handle online teaching			2-c
Attitudes toward teaching and life			2-d
Knowledge of assessment theory			2-e
Identity/Position			2-f
Assessment changes	Planned changes	Summative assessment	3-a-i
		Formative assessment	3-a-ii
	Improvised changes	Summative assessment	3-b-i
		Formative assessment	3-b-ii

that making some adaptations to meet the needs of this new way of teaching might be useful, he felt some challenges imposed by online teaching and did not make many changes to his assessment practices, as illustrated below:

[Excerpt 1].

I have taught for over 20 years. To be honest, I have made few changes to my way of teaching...I am a person that doesn't like change...Teaching online was totally a new thing to me. Familiarizing myself with and learning how to use the online teaching software was already a huge burden for me... I don't have extra energy to think about how to change the way of assessment. My goal for last semester was to complete my teaching tasks.

Instead, he offered a make-up, paper-and-pencil final exam at the beginning of Fall 2020 (when the students returned to campus) for his spring course so that he did have to think about ways of changing his assessment practice. When asked what he would do if the pandemic was not under control and the lockdown was not eased in Fall 2020, he admitted that he "didn't think much about that" and planned to "take one step at a time and wait and see." It can be seen that Jim's own motivation, the ability to adapt to new environment, and the option of making up for the exam in Fall 2020 offered by the university collectively influenced his way of assessment.

Helen was similar with Jim to have only raised the ratio of formative assessment and not make any other changes to formative assessment. However, her rationale was different.

[Excerpt 2].

I only had 10 students in my class. With this small class size, I could ask them to have their cameras on and the internet connection was fairly smooth. I could see their faces.... For me, it was just moving the classroom online. I used Tencent Meeting to lecture, and I lectured the same way as I did in the classroom.

When we asked whether she was concerned about the quality of teaching and learning in an online environment, Helen commented:

[Excerpt 3].

The pandemic has impacted all people's life, work, and travel. Definitely, it would influence teaching as well. The whole country was locked down and everyone had to stay at home. It was an extremely difficult time for all. I think under this situation, we should be cautious against being too demanding on the students. As long as they could learn something, it should be fine. It's already a fortune if they stay healthy physically and mentally.

It can be seen that the class size and Helen's attitude toward teaching during the pandemic were the main factors leading to why she did not change much of her formative assessment. As for the summative assessment, she, after asking students' opinions, turned the end-of-semester group project to individual work since students complained about the ineffectiveness of group collaboration.

Cathy and Henry

Cathy and Henry both made changes to their assessment practices as their classes moved online. As they commented, these changes were made to ensure test fairness for the students. However, the changes were mostly made on the format to accommodate the online instructional and assessment modality. Cathy made changes to formative assessment but not summative assessment because the university provided the option of having students make up for the final exam if teachers applied for this option. She raised the ratio of formative assessment from 20 to 50%, included more in-class discussion and after-class written homework, and increased the use of peer-review for essay writing. In her interview, she explained that raising the percentage of formative assessment was a mutual decision at a meeting among her colleagues. Cathy commented that the challenge in teaching and assessment online is largely due to the internet connection. Especially with a large class, she could not ask students to turn on their cameras or easily check whether students were paying attention during class. To assess what the students were doing, she “asked students questions frequently” so that they did not “get distracted by other stuff such as TikTok videos and computer games.”

Cathy also increased the use of peer-review. She did not plan to do this, but since she asked students to do peer-review, she found that her students provided more feedback to each other's essay and the comments were more specific, straightforward, and constructive compared to face-to-face peer-review. This might be because, according to her, “students could not see each other, and they did not need to worry if they ‘threatened other’s face’.”

Henry also made changes to meet the needs of online teaching. He raised the ratio of formative assessment (from 30 to 60%), stopped using quizzes, and added more discussion and written homework. Henry stopped using quizzes because “students could easily find the answer keys online, and it might be unfair to those who did not ‘cheat’.” Instead, he included more written homework and added more in-class discussion by asking students to type their answers in the chat box of the online teaching platform to check if students understood his lecture. As illustrated below, Henry did so because.

[Excerpt 4].

In classroom, when I saw puzzled faces, I knew they had questions and I would provide more explanation through repeating or rephrasing. However, when teaching online, I was not able to see them.... Asking them all to turn on their cameras was not feasible because the internet connection would get really bad and my students wouldn't hear me clearly.

To solve this problem, he asked all students to type their answers in the chat box after he asked a question. By doing this, he was able to diagnose if students had questions and interact with them to better address their questions.

As for the summative assessment, he turned the final exam from a closed-book, timed writing test to open-book, untimed writing. Henry made such a change given that the poor internet connection would make it difficult for him to proctor for a closed-book test online and prevent students from cheating (e.g., referring to notes or searching information online during the test). When we asked whether he attempted any online proctoring services or tools, it seemed clear that Henry did not explore potential proctoring services, and he explained:

[Excerpt 5].

I am a tenure-track, research fellow ('Qingjiao'). My employment contract requires a large number of publications. As a 'Qingjiao', I am under a lot of pressure to publish, so I need to focus on research, and not teaching.

Although he did not incorporate technology into online assessment, he believed he had “tried his best to ensure the effectiveness and fairness of assessment.” He also expressed his willingness to reform assessment had he had less pressure from publishing, stating that “had I been tenured, I could have devoted more time to teaching reform and might have explored technologies that could facilitate online assessment.”

Mary and Steve

Mary and Steve made the most changes to their assessment practices to meet the needs of online teaching and assessment during the pandemic. They did not just change the assessment format; instead they made more fundamental changes to align assessment, teaching, and learning in the online environment. Mary made changes to both formative and summative assessment for her translation course. She increased the ratio of formative assessment, assigned more translation homework, and used the learning management system to track students' participation in class. She also removed the in-class group work due to the constraints of online teaching. As she reasoned below,

[Excerpt 6].

When teaching in class, I could assign students into groups, and they could immediately start discussion and group work, and I could walk around to see how each group was doing. But when teaching online, after I assigned them into different groups, they had to log out from the teaching platform and start another group meeting using a different application. I

could not join their discussion and check what they were doing, either. It was troublesome, time-wasting and ineffective, so after a few times of group work, I abandoned it and turned it to individual work.

Similar to what Henry did to his summative assessment, Mary also turned the closed-book test to an open-book test, arguing that the open-book test was more appropriate for online assessment and could be fairer to all students. However, different from Henry, Mary also changed the rubric, stating that:

[Excerpt 7].

Before, I focused more on the correctness and completeness of the translation. With more references to turn to, almost all students could meet the requirement 'correctness and completeness of translation'. Thus, I added idiomaticity in the target language as a new criterion.

During the interview, we also learned that Mary had some knowledge of assessment theory as she reported that she took a language testing course when she was doing her PhD study, so we asked her whether the course could help with the adjustment in online assessment practices, and she commented, "It has been a long time. I almost forgot what I learned, but I knew I needed to adjust rubric to different ways of assessment. So, yes, the assessment knowledge did help."

Among all the participants, Steve made the biggest changes to his way of teaching and assessment. During the pandemic, he carried out blended learning by using a massive online open course (MOOC) that matched his course to set up a small private online course (SPOC) for his own students. Most of his assessments were done on the SPOC platform. He asked students to watch a video every week and complete a quiz on the content of the video. Students must complete this assignment before the online, synchronous session every week. After each session, he would ask students to do a written-assignment or participated in an online discussion regarding the content of his lecture so that he could evaluate their class performance. These were also done on the SPOC platform. He did not use blended teaching before the pandemic, and it was the first time for him to try such way of teaching. Steve explained,

[Excerpt 8].

To be honest, I had always been thinking of reforming my way of teaching and assessment, but the pandemic motivated me to do so. In addition, our university had been encouraging teachers to carry out blended teaching even prior to the pandemic. I like to explore new technology. The university invited many professionals who have experience in online teaching

and blended learning to give lectures. Plus, during the pandemic, many massive online open courses were free and available to all teachers in China, so I could use other teachers' online course to create my own SPOC.

When it comes to assessment practices, he also raised the percentage of formative assessment and told students, at the beginning of the semester, the detailed grading breakdowns of different types of formative assessments (e.g., how many quizzes and how much each quiz constitutes, how many times of written homework, etc.)." He further explained,

[Excerpt 9].

with the SPOC platform recording all students' record and with all quizzes prepared by other teachers before the course began, I was able to tell my students the detailed grading breakdowns of the formative assessment and used the results to inform my teaching in a better way.

Steve perceived using SPOC to carry out blended teaching and online assessment as effective, and he decided to "continue to use such way of teaching and assessment in Fall 2020 when students came back to campus.

Discussion

The interview data with EFL teachers at University A showed clear changes in their online assessment practices during COVID-19. Despite the national and local (universities') policies on the shift from face-to-face to online instruction, the lack of time and prior experience in online teaching and assessment led to a great degree of variability in EFL teachers' online assessment practices. The results suggest that the adjustments in the assessment practices were a mixture of planned and improvised changes. This is inevitable as classroom scenarios were dynamic, and teachers often "improvise in their assessment practices, besides what they have planned beforehand" (Erickson 2010, cited from Xu, 2016, p. 2017). Despite the wide variety of changes the EFL teachers made to their assessment practices, those changes were not made arbitrarily. Rather, their decision to make changes (or not) to their assessment practices was mediated by both contextual factors and experiential mediators. Moreover, the contextual factors prompted the teachers to make changes to their assessment practices; however, it was the experiential factors that brought out more individual differences in teachers' online assessment practices. Therefore, we will discuss the results from two aspects: improvised and planned changes, and contextual and experiential factors.

Planned and Improvised Changes

Some teachers' changes of assessment practice were made before the beginning of online teaching and some of the changes were made in the ongoing process of assessment.

Some planned changes were initiated by the university, e.g., raising the ratio of formative assessment and replacing close-book exams with open-book ones. Some changes were planned by teachers, e.g., Steve planned to use SPOC to conduct the assessment and clarified the breakdowns of formative assessment. Henry stopped using quizzes since he thought of students' possibility to cheat at the beginning of the semester; quite a few teachers included more written and translation homework as alternative assessment. Cathy included more in-class questions and answers to assess whether students were paying attention.

As classroom scenarios are fluid and dynamic, teachers have to adjust their assessment practice, and therefore, improvised changes were made. Some improvised changes were initiated by others. For example, Helen changed the final group work to individual work because her students told her about the difficulty in conducting online group work. Some improvised changes were initiated by teachers themselves as they were assessing and reflecting, and their reflections influenced their assessment so that they made changes in the ongoing process. For example, Mary noticed the ineffectiveness of in-class group discussion after using it a couple of times, then she stopped using it and replaced it with more individual work. Other self-initiated changes included the use of chat-box discussion and peer-review. As Henry found the efficiency of having students type in the chat-box on the teaching platform in class, he increased this way of in-class assessment. Similarly, Cathy found that students did a better job in on-line peer-review than in class, so she used more peer-review.

The planned changes tended to be a result of top-down policies which were planned ahead of time by the teachers. However, in the process of teaching and assessment, as teachers reflected on their assessment practice, they also adjusted their ways of assessment in a bottom-up fashion to meet the needs of online assessment to achieve the best teaching and assessment results. This aligns with the findings in Xu and Liu (2009) in that "teachers need space and resources to adjust their assessment practice (p. 509)." Provided with enough space and resources, teachers can learn about assessment on the job.

Contextual Factors

The contextual factors included the national context and the local context. At the outbreak of the pandemic, the Ministry of Education initiated the top-down policy of teaching online, which was also a national contextual

factor. This national context is the driving force for all the changes in resources and practices related to online teaching and assessment across contexts. Under this backdrop, more online courses were developed and made available to teachers and students. An example of this impact, as revealed in the interview data in this study, is that iCourse, China's largest online learning platform, created over 8,000 courses freely available to teachers and students, so that teachers can utilize these resources to make changes to their online teaching and assessment practices.

Aside from the national policy, the local context played a more important role in mediating teachers' specific online assessment practice. At the institutional level, schools create context-specific policies and resources to comply with the national policy to accommodate teaching and assessment for the COVID situation. Specifically, University A asked teachers to raise the ratio of formative assessment, encouraged the use of open-book exams, and offered the option of make-up exams when students returned to campus. The university also called for the use of blended teaching and invited experts and experienced teachers to give lectures on how to carry out online teaching and blended learning. Meanwhile, the university was flexible as to which platform and what forms of assessment teachers decided to use. These factors all influenced teachers' decision in making changes or not to their assessment practices.

To comply with the national and local policies, teachers also need to consider other practical constraints when revising their online teaching and assessment practices, which include, but not limited to, class size, internet connection, and effectiveness of technology. For small-size classes (e.g., Helen's class), the teacher could have students turn on their cameras and observe them face to face; however, for large-size classes, teachers could not ask all students to turn on their cameras since the internet connection would be bad, so teachers had to think of other ways to assess students' class performance. Technology also constrained group work: without a platform or application that could host group work available to Chinese EFL teachers, they had to cut off group work.

Taken together, these contextual factors provided a larger direction of change for teachers, identifying the domains and range of assessment activities teachers could adopt in their online assessment practices. However, these policies and resources were also broad enough to leave sufficient freedom and flexibility in teachers' individual decisions on their assessment practices. Such flexibility led to a noticeable level of variability in teachers' assessment practices.

Experiential Factors

The individual differences in teachers' assessment practices can be further explained by experiential factors. Such experiential factors included teachers' motivation, attitudes toward new assessment methods, ability to handle online teaching and technology, attitudes toward teaching and life, and knowledge of assessment theory. It was found that teachers who had higher motivation and an open mind toward new assessment methods tended to make more changes to meet the needs of online assessment. However, teachers who were reluctant to change made fewer changes. For example, Steve, who showed strong motivation to reform, had thought of reforming his own assessment practice even before the pandemic broke out; and the changes he made to the assessment activities were also more substantial than other EFL teachers we interviewed. In contrast, Jim, who did not show a strong motivation, only changed his assessment practices minimally to comply with the national and local policies.

Another experiential factor pertained to teachers' knowledge of assessment theory. Those who had some assessment knowledge and background (e.g., Mary and Steve) made more substantial but thoughtful changes to their assessment practices. For example, Steve used a brand-new mode to meet the needs of online assessment by making use of SPOC platform, because he recognized the importance of alignment between teaching, learning and assessment in the online environment. To him, when teaching and learning change substantially in the online environment, assessment must follow. In another example, when the original face-to-face final exam was converted to an open-book, online exam, Mary also changed her rubric to align with the online delivery format and any influence this format might have on test performance. Note that we are not arguing that teachers without assessment knowledge and background made inappropriate or random changes. Rather, our observation in this study showed that the EFL teachers less trained in language assessment tended to make changes based on their teaching experience more often than considerations of assessment concepts and principles. That said, most teachers recognized the potential issues related to fairness when making changes to their assessment practices.

A third experiential factor was teachers' attitudes toward assessment, in particular the role of assessment in learning and even life of the students. For some teachers (e.g., Helen), the pandemic was a difficult situation for everyone. From their perspectives, too many changes to the assessments might unnecessarily cause excess level of stress on the students, who were already struggling to keep up with learning during the pandemic. Therefore, while they spent much effort ensuring the quality of online instruction, they

did not make many changes to their assessments, and their goal of assessment was to let students pass the course.

The final experiential factor related to identity. In this study, it appears that the EFL teachers adopted different identities that were associated with different responsibilities and priorities at their jobs; and their identities seem to be associated with their perceptions about language assessment and the role of assessment in their profession. Both Steve and Henry were young lecturers, but they displayed different practices in terms of using technology for online assessment. Steve's position was a permanent one while Henry's was a tenure-track position. As an early-career researcher (ECR), Henry was under the pressure of "publish or perish" and ECRs ("Qingjiao" in Chinese) in China are also facing the challenge of balancing research and teaching (Hu, 2015). It can be seen that some young Chinese lecturers' identity as an early-career researchers and huge pressure of publishing they were facing may have influenced their practice of online assessment during the pandemic.

Conclusions, Limitations, and Implications

In this study, we investigated Chinese EFL teachers' online assessment practices during the pandemic, in particular, how those practices differed from their traditional in-person assessment practices and what caused those changes. Findings of this study provide a glimpse of the online assessment practices of EFL teachers in a Chinese university as well as the mediating factors that influenced their assessment practices. The EFL teachers made assessment decisions and selected assessment methods based on policy, the local context, and their own teaching experience and reflections.

This study is limited in its small sample size. We only investigated online assessment practices of six EFL teachers at one university, and therefore, the results of the study could not be generalized and should be interpreted with caution. Future studies can adopt different methods, e.g., large-scale survey, with a large number of participants in different levels of universities across different regions in China to add the representativeness and generalizability of the research.

Despite the aforementioned limitations, this study has important implications. First, this study shows that language teachers' online assessment practices are fluid and context-dependent. Teachers made both top-down planned changes and bottom-up improvised changes, and such changes are mediated by both contextual and experiential factors. Therefore, teachers, when making changes to adjust to different assessment needs, should plan for changes ahead of time, but also be ready to face

unexpected problems and make improvised changes to address those problems such as students' needs, classroom environment, internet connection, and teaching resources. Second, in alignment with previous literature (Xu & Liu, 2009) which emphasized the importance of teachers' autonomy in assessment, this study also demonstrates that when given autonomy by the government and university, teachers are able to make their own assessment decisions and adjustments to meet the needs of new teaching environment. Therefore, when implementing online assessment policies or providing online assessment training, policy makers and assessment trainers should consider teachers' autonomy and agency. Future research can also focus on teacher agency to obtain a fuller picture of teachers' online assessment literacy.

Acknowledgements This project, undertaken by Cong Zhang, is funded by Shandong Provincial Social Science Foundation of China (Grant No.20CWZJ28). We are grateful to the teachers who participated in this study. Without their support, this study would not have been accomplished.

References

- Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349–383.
- Bryman, A. (2015). *Social research methods*. Oxford University Press.
- Crusan, D., Plakans, L., & Gebriel, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43–56.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25, 327–348.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132.
- Harper, C., Platt, E., Naranjo, C., Boynton, S., Ramanathan, V., & Morgan, B. (2007). Marching in unison: Florida ESL teachers and no child left behind. *TESOL Quarterly*, 41(3), 642–651.
- Hu, Z. (2015). The dilemma of Qingjiao's growing and its solution. *Contemporary Youth Research*, 6, 54–58.
- Jin, Y., Wu, Z., Alderson, C., & Song, W. (2017). Developing the China Standards of English: challenges at macropolitical and micropolitical levels. *Language Testing in Asia*, 7(1), 1.
- Kleinsasser, R. (2005). Transforming a post-graduate level assessment course: A second language teacher educator's narrative. *Prospect*, 20, 77–102.
- Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120(2), 285–285.
- McNamara, T. (2011). Managing learning: Authority and language assessment. *Language Teaching*, 44(4), 500–515.
- North, B. (2000). *The development of a common framework scale of language proficiency*. Peter Lang.
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381–402.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429–462.
- Reynolds-Keefers, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research and Evaluation*, 15(8), 1–9.
- Rohl, M. (1999). Profiling ESL children: How teachers interpret and use national and state assessment frameworks. *Queensland Journal of Educational Research*, 15(1), 113–122.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402.
- 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., & 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.
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teachers' experience. *TESOL Quarterly*, 43(3), 493–513.
- Yan, X., & Fan, J. (2020). "Am I qualified to be a language tester?": Understanding the development of assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246.
- Yan, X., Zhang, C., & Fan, J. (2018). "Assessment knowledge is important, but...": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158–168.

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