Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Covid-19 as a critical incident: Reflection on language assessment literacy and the need for radical changes

Wenwen Tian a, *, Stephen Louw b, Muhammad Kamal Khan c

a School of Foreign Studies, Northwestern Polytechnical University, Xi’an, Shaanxi, China
b Department of Language Studies, School of Liberal Arts, King Mongkut’s University of Technology Thonburi, Bangkok, Thailand
c Department of English and Applied Linguistics, Allama Iqbal Open University, Sector H-8, Islamabad, 44000, Pakistan

ARTICLE INFO

Keywords:
Language assessment literacy
Critical incident
Duoethnography
Reflective practice
Teacher professional development
Covid-19

ABSTRACT

Three participant-researchers explored their language assessment literacy (LAL) in the context of school shutdowns during the Covid-19 pandemic in three Asian countries: China, Thailand, and Pakistan. We employed a duoethnographic approach to explore our understandings of LAL, our experiences in language assessment practices, and our perspectives on our professional development during the disruption. Data were collected in January and February 2021, a point at which the school shutdowns had become a reality for teachers across Asia. Data included written narrative reflections and Zoom discussions framed as reflection-on-action, reflection-in-action, and reflection-for-action, focusing specifically on assessment issues. Our initial reflections show a lack of confidence in our LAL and a perception of assessment as highly theoretical. As the study progressed, our reflections on assessment become more confident with increasingly sensitive topics, including the value of training and our role as assessment practitioners in the hierarchical bureaucracy. This study highlights that remote reflections with colleagues across cultures and contexts may be highly constructive in promoting professional development. It is recommended that front-line teachers be given opportunities to develop their own LAL through interfacing with other professionals through reflective practice within and beyond their cultural contexts.

1. Introduction

The Covid-19 pandemic and the resulting school shutdowns have triggered a tremendous shift in educational instruction as teaching and learning move online, creating a variety of challenges for educators worldwide (Bozkurt & Sharma, 2020; Zhang, Yan, & Wang, 2021). Assessment of teaching outcomes and students’ performance has been made particularly challenging since neither on-site large-scale summative assessment nor conventional formative assessment can happen as normal (UNESCO, 2020). Under these conditions, teachers face an imperative to embrace pandemic pedagogies, digital literacy, and alternative approaches to assessment which challenges traditional conceptualizations of language assessment and conventional test delivery modes.

In this paper, we argue that in responding to the disruptive and stressful demands of pandemic-valid assessments, there is potential for meaningful professional development of teachers’ approaches to, understanding of, and practices in assessment. We believe that a crucial factor in facilitating developmental benefits for teachers from the disruption to routines, such as the Covid-19 pandemic, is the mobilizing of reflection, and particularly dialogic reflections with peers. Reflection forms a critical dimension of teachers’ professional
development. It can assist teachers in improving their language assessment practices and promoting their LAL (Babaii & Asadnia, 2019). To this end, we employed a duoethnographic approach to engaging in reflection on our language assessment practices during the school lockdowns. Through this study, we hope to shed some light on how teachers’ reflection on critical incidents can be instrumental for professional development in terms of LAL.

2. Theoretical underpinnings

2.1. Conceptualizing language assessment literacy in language education

Since Stiggins (1991) affirmed that “we are a nation of assessment illiterates” (p.535), the significance of assessment literacy has been a focus of scholarly attention. In the field of language teaching, LAL refers to the knowledge and skills stakeholders need to conduct language assessment activities where stakeholders may include classroom teachers, test writers, institutional administrators, and professional language testers (Taylor, 2013). A teacher’s LAL is conceived as a complex multi-layered enterprise needed for language assessment, including a repertoire of knowledge (language pedagogy, theory and concepts, own language assessment context), skills (instructional skills, design skills for language assessments, skills in educational measurement, technological skills), and principles (awareness of and actions towards critical issues, such as sociocultural values, local practices, and personal beliefs/attitudes) (Giraldo, 2019; Taylor, 2013). However, according to Scarnino (2013), knowledge, skills, and assessment principles are insufficient for language-assessors as they go about performing assessment since LAL also requires:

interpretive frameworks which are shaped through their particular situated personal experiences, knowledge, understanding and beliefs. They are obliged to integrate simultaneously the complex theoretical, practical and institutional dimensions of the assessment act and an understanding of the interpretive nature of assessment. (p.322)

This interpretive framework is a key to LAL, but how can teachers go about developing it? Formal training of language assessment in degree programs or in-service workshops is a natural starting point. Training can introduce essential aspects of theory and practice with which teachers can develop a yardstick to measure their skills. Scarnino (2013) advocates training teacher-assessors to explore and evaluate their own preconceptions in order to interpret their own assessment practices and their students’ learning. She believes that through these processes teachers gain critical understanding and self-awareness of the interpretive nature of assessment.

However, there are two obstacles which stand in the way of training teachers with this self-awareness. The first is the daunting range of knowledge and skills required as part of developing their LAL (Taylor, 2009). Comprehensive formal training on the full spectrum of theories and principles in assessment can bewilder a beginner, and overwhelm the already busy classroom teachers. Despite the pervasive rhetoric in recent decades about the importance of assessment in the teaching and learning processes, teachers still report a lack knowledge, experience, and confidence to perform language assessment (Herrera & Macías, 2015; Hidri, 2021). The second challenge to the LAL development is more systemic. Previously, testing or assessment were viewed and practiced using a top-down approach, in which the competencies of measurement would trickle down from “assessment literate” professionals to the assessment illiterate teachers below (e.g., Stiggins, 1995). This hierarchical test-centric paradigm is still evident in Asian countries, including China (Jin, 2010), Thailand (Watson Todd, 2019), and Pakistan (Ashraf & Zaki, 2019), where we may find individual teachers have less incentive but practical needs to developing their LAL (Coombe, Vafadar, & Mohebbi, 2020). If language assessment is hierarchically controlled, there is arguably little need for an equal distribution of assessment literacy across stakeholders. In contexts with democratized assessment (Inbar-Lourie, 2008), practice and theoretical knowledge of assessment can be shared across teachers and other literate stakeholders.

Challenges such as these notwithstanding, the development of LAL at the teacher level is crucial. We agree with Scarnino (2013; 2017) that each teacher needs their own interpretive framework for their LAL, one that is relevant to their contexts and the stakeholders with whom they work, and is born from their personal experience. Developing LAL necessarily requires opportunities for teachers to engage in making decisions about assessments that are appropriate for their own -classroom contexts. Crucial of this is the role of teacher reflection, and under the right conditions, the serendipitous role of critical incidents as a catalyst for reflective engagement.

2.2. Critical incident as a reflective tool for professional development in teacher development

In education, a critical incident is any unplanned and unanticipated event that occurs during class, outside class or during a teacher’s career but is “vividly remembered” (Brookfield, 1990, p.84). These critical incidents are described as events that interrupt (or highlight) “the taken for granted ways of thinking about teaching” (Farrell, 2008, p. 3). While Farrell’s (2008) definition of critical incidents refer to local events at the classroom level, Tripp (1993) takes a broader, historical view in which a critical incident “refers to some event or situation which marked a significant turning-point” (p. 24). The Covid-19 is such an event. It has marked a significant turning point for teachers and educational system globally. Since the start of the pandemic, there has been a major reorientation towards technology as the basis for delivering tests and making assessment. For those who reflect formally on their pedagogical practices by articulating and analyzing this new set of demands, this is an opportunity for professional development. In this study, we approach the Covid-19 pandemic as a critical incident which has impacted our language assessment practices in our respective sociocultural contexts.

Critical incidents form a valuable source of professional development for teachers (Joshi, 2018). In teacher development, critical incidents and the resulting reflection serve to raise teachers’ self-awareness and encourage teachers to ask critical questions about
established routines and procedures (Richards & Farrell, 2010; Yu, 2018). A key issue with the critical incidents as a form of pedagogical development for teachers is that they are matched to a meaningful reflective response to the incident for transfer to classroom practice (Farrell, 2013). Following Schön (1991) and van Manen (1991), we conceptualize reflection as three interrelatedchronological categories: 1) reflection-on-action involves thinking back on previous experiences and evaluating possible alternatives for improvements; 2) reflection-in-action is active or interactive reflection which involves examining how beliefs and experiences are connected to theories-in-use; and 3) whereas reflection-on-action and in-action relate to teachers’ past and present exigencies, reflection-for-action enables practitioners to evaluate their strategies or techniques to enhance their practice for future development.

Crucially for the current study, reflecting collaboratively plays a key role in expanding one’s perspectives through sharing and interacting with teachers in other contexts (Abrahamson & Chase, 2015). In sociocultural theories of teacher development, meaning-making is situated through meaningful social interactions (Lantolf & Johnson, 2007). Here, reflective activity is not simply individual internal exploration, but social interaction with a peer which leads to the co-construction of new knowledge, perspectives or beliefs. It is this collaborative reflective practice which we apply in this study to reflect on our LAL.

This study is set within the massive disruption to the educational institutions caused by the Covid-19 pandemic. During the initial stages of the pandemic, our focus as educators had been on the challenges of dealing with teaching using online platforms. By the end of 2020, the demands for a solution to assessing students remotely had become a reality. With this as a driving context, we embarked on an investigation into our response to these challenges under the theoretical framework of LAL.

3. Methodology

3.1. Participants

The authors were participant-researchers in this study. We each hold a PhD in Applied Linguistics and have extensive experience as English language teachers, but are not specialists in language assessment. Therefore, we each approached this study as ‘knowledgeable beginners’ (our self-perception) in language assessment and from our respective perspectives in Asia: Wenwen in China, Steve in Thailand, and Kamal in Pakistan.

Wenwen completed her undergraduate studies and initially worked as an English language teacher in China. She then earned her MA and PhD in Thailand, where she also worked as an English language teacher, and then worked for international affairs at a university in Saudi Arabia for three years and spent one year in the USA where she obtained two certifications: TEFL (Teaching English as a Foreign Language) and CELTA (Certificate in Teaching English to Speakers of Other Languages). In total, she has 14 years’ experience as a language teacher. Her experience abroad brought to her attention the pedagogical and cultural practices in Chinese education she had not questioned before. Where China’s educational systems are highly centralized (Huang, Wang, & Li, 2015), she found that in Thailand teachers have greater decentralization to implement courses and assessments, and are therefore less bound by bureaucratically prescribed outcomes.

Kamal works at a large national university in Islamabad, Pakistan. His area of expertise is phonology and he holds a Post Graduate Diploma in ELT. At the time of the study, he had been teaching English for 18 years. Pakistan’s educational system is largely centralized and highly bureaucratic (Coffman, 2005). Although his progress as an educator has largely taken him through the Pakistani educational system, Kamal has experience beyond Pakistan through his involvement in workshops and short courses in the USA, UK, and Thailand.

Steve, originally from South Africa, has a career spanning over 30 years as a language teacher. Being specialized in teacher training, he works mostly in private language centers in Singapore, Malaysia, Thailand, and Cambodia. His language teaching experience has differed markedly from that of Wenwen and Kamal. As a ‘foreign’ teacher in private schools outside of the official government teaching systems, he has not been subject to top-down bureaucratic demands, and has worked most often with small classes which have allowed for opportunities to implement communicative teaching methods.

Although qualified and have experience in language teaching, we each considered LAL to be a new field of inquiry, and were inexperienced in the remote assessment that resulted from the pandemic school lockdowns. Based on these unique challenges these posed, we aimed to explore our responses to the question of how to manage language assessment in our respective contexts.

3.2. Research approach: duoethnography

Given the geographic distance between us, we framed our reflections on our LAL within the Covid-19 pandemic as a duoethnography. With its roots in ethnography, duoethnographic research involves two (or more) individuals who foreground themselves as sites of research by engaging in a dialogical inquiry into a shared social phenomenon (Norris & Sawyer, 2012a). Duoethnography allows researchers to construct situated and thick descriptions of their own lived experiences through dialogue-based accounts of how they make sense of the (re)creation of their experience within “enculturating discourses” (Norris & Sawyer, 2012b, p. 290). Data in duoethnographic research may take forms of conversation (Lowe & Kiczkwia, 2016) or extracts of narratives (Rose & Montakantiwong, 2018) to make individual voices explicit and distinctive on a shared phenomenon (Banegasa & Gerlachb, 2021).

In English language education, duoethnography is increasingly used as a means of exploring the inner world of teachers. For instance, Banegasa and Gerlachb (2021) use a duoethnography to uncover teacher educators’ identities and agency, Sebok and Woods (2016) look at teacher educators’ understanding of professionalism, and Rose and Montakantiwong (2018) illustrate the challenges of teaching English as an international language. The appeal of duoethnography for us is that it creates a space to explore a common topic from our disparate contexts, and thereby reconceptualize and rethink topics around LAL dialogically in collaborative reflection.
3.3. Data collection

Data were collected in January and February 2021, a point at which the impact of the pandemic on educational systems had become evident. The data took two forms: written reflections and online discussions. Because of our geographical distance, we felt that the written medium to record our narratives would serve us best, allowing for time zone differences and giving us time to carefully consider one another’s contributions before commenting. We opted to share our reflections as a Google document and supplemented these written narratives with online discussions using Zoom, which would help build greater rapport. It also allowed us to unpack issues and go into topics arising in the written reflections.

The written reflections took the form of three separate documents. Our first reflection was an opportunity to get to know one another in the context of our professional work with language assessment. The early turns of this reflection established our respective work experience and the training related to testing and assessment. The data collection proceeded and we became more confident with the medium of remote reflection, our conversations approached increasingly sensitive experiences we wrote about, and elaborate on the ideas presented in the texts.

Following each narrative, our Zoom meetings served to collectively interrogate the issues raised in our written reflections. Remote meetings such as these have been used elsewhere to facilitate reflection, increase a sense of community, and engender professional growth (Burhan-Horasanli & Ortaçtepe, 2016). Each meeting lasted around one hour, and was conducted after each of us made contributions to the written reflections. Since the meetings also served as an opportunity to meet one another, much of the initial and final turns of these meetings were taken up with off-topic social exchanges about issues unrelated to LAL, which we ignored in the data analysis. In our on-task dialogues, we were able to clarify the ideas expressed in our written reflections, provide greater context for the experiences we wrote about, and elaborate on the ideas presented in the texts.

3.4. Data analysis

We took the three written narratives as the primary source of data and the transcripts from three Zoom meetings as supplementary. The reflections totaled 12,766 words (5,450 words in the first, 3,958 words in the second, and 3,358 words in the third reflection). The three Zoom meetings generated three-hour-and-fifty-minute data, which were transcribed, producing a total of 30,789 words (11,422 in the first, 10,912 in the second, and 8,455 in the third). Although we aimed to have a specific focus for each reflection (for example, reflection-for-action in our third round), we became sidetracked, indulged in anecdotes, and found topics that elicited highly emotional responses. Although rich, the range of topics was far broader than we had anticipated and proved challenging to analyze.

To prepare the data for analysis, each document was labeled and color coded to indicate speaker and dates. With Farrell’s (2012) framework as a starting point, we explored the range of topics and issues covered in the data using Cohen, Manion, and Morrison’s (2018) principles of breaking down, examining, categorising and reconstituting data in new ways to make connections between different aspects of the complete data set. We each coded the data separately, looking for emergent and recurrent topics within the data, then met on Zoom to compare findings and agreed on the unifying themes to report.

Before we proceed to the findings, we wish to identify weaknesses in our data and interpretation. Firstly, as the data collection proceeded and we became more confident with the medium of remote reflection, our conversations approached increasingly sensitive topics, especially as we began to embraced more emotive issues, and tended to stray off-task. In terms of the study, this proved interesting, but as will become evident, we did not achieve our research goals. Secondly, as expected from ethnographic data, our findings represent an emic perspective (Blommaert & Dong, 2020) based on our contexts and experience, and reflect our personal understanding of the interaction between the three contexts and the three stages at work (before, during, and after Covid-19). As a result, no claims of generalizability can be made. To increase transparency and trustworthiness in our analysis, in what follows we present lengthy extracts as far as space allows to offer the reader a meaningful window into our individual contexts and our interpretation of the data.

4. Findings

4.1. Reflection 1: Our LAL before Covid-19

Our first reflection was an opportunity to get to know one another in the context of our professional work with language assessment. The early turns of this reflection established our respective work experience and the training related to testing and assessment. The following excerpt is a sequence of turns in which Steve and Kamal respond to Wenwen’s contribution.

Wenwen (Reflection 1 turn 15): Looking back to my teaching career path, I have been working as an EFL teacher at universities in China and Thailand for nearly 14 years since I graduated with a Bachelor of Art Degree [in language studies] in northwestern China. Looking back to my teaching path, I have to say that I have never been trained explicitly to do language assessment. I should have gained some knowledge and skills in designing formal/standard tests and alternative assessments from years of experience through classroom teaching in China and Thailand. But the number of years of teaching experience does not necessarily translate into expertise.
Steve (Reflection 1 turn 16): Yes, me too! I’ve been teaching for 30 years now, and I’ve sort of actively avoided the whole question. I prefer collecting scores from my students through informal means or through activities that lead to formative feedback because I know that my test-making skills are weak. I have made language tests, of course, but I am well aware they are not good. They are time consuming to make, and in my experience here in Thailand the schools always have a lot of criticisms about what you give to the students.

Kamal (Reflection 1 turn 17): In retrospect, when I look back at my early career assessment and evaluation strategies and techniques (that I used as a teacher) I feel that things were more theoretical than practical. We were not properly given chances to see the applied aspects of [language assessment]. We rarely applied our critical thinking and creativity in assessing and evaluating our students. Until Covid-19, we never used technology for assessment and evaluation except those very limited usages e.g., similarity checking and reviewing papers.

This sequence highlights various themes which emerged from the first reflection. The first of these is the repeated reference to the various influences which have informed our LAL. We each make it clear that we are drawing from a respectable well of experience as teachers. This experience forms an important platform from which we generate our views on LAL. In earlier turns, we referred to our experiences as school students, which Wenwen (turn 12) describes as ‘exam-oriented and test-driven’ and Kamal (turn 13) notes as ‘product oriented’ in which ‘our focus remained all the time achieving good grades’. Such early experience of testing forms part of our apprenticeship of observation (Lortie, 1975) establishing early expectations of assessment practice (Xu, 2019).

Given our extensive experience and qualifications, there is reason to believe we would be speaking about our language assessment from a position of knowledge. However, there are repeated occasions across the data in which we call attention to the inadequacy of our LAL. In the extract, Steve judges his test-taking skills as ‘weak’, and that the language tests he has made are ‘not good’. Disparaging comments about our perceived LAL occur in 19 turns across the data. In our analysis, we considered the possibility that these self-deprecating comments were a form of false modesty (Murray, 1982), but it is more likely that they reflect the reality of a lack of self-confidence in our LAL. This finding is consistent with other work (e.g., Tsagari & Vogt, 2017) highlighting teachers’ low confidence in their LAL.

One explanation for our lack of confidence is the mismatch we experience between the theory and practice of language assessment. Kamal’s description of ‘not properly given a chance to see the applied aspects of language assessment’ highlights his perspective of a separation of the theory of language assessment and its application, and that his training was ‘more theoretical than practical’. Wenwen’s claim that she has not been ‘trained explicitly to do language assessment’ probably reflects this same praxis disjunct; ‘explicitly’ here indicating her belief that any theoretical perspectives she encountered were somehow removed from the practice of language assessment. In this context of the separation between theory and practice, it is worth scrutinizing Steve’s dismissive comment that assessment can be ‘actively avoided’. Steve expresses a preference for ‘informal’ assessment, which he equates with formative feedback, and which he presumably distinguishes from ‘formal’ assessment (to be avoided). Our reflections on our LAL, then, show how we have each somehow distinguished between the practice and theory of assessment, on the latter have found ourselves lacking.

The low confidence in our LAL and the theory-practice disjunct lead to the third theme identified in this reflection: our view of the complexity of language assessment. Kamal, for instance, describes language assessment as requiring ‘critical thinking and creativity’. This sense of language assessment as a highly complex and specialised field demanding specialist jargon that is beyond the capabilities (or needs) of a normal classroom teacher is developed by Steve (Reflection1, turn 35):

I think another reason why I’m not good at formal tests is that during my undergraduate years I did a course in psychology on psychometric tests, and we covered all the ways that tests create bias and how to avoid it. There were long lists of ways tests can go wrong, and long lists of how to prevent them - using positive and negative test forms, ways of ensuring validity, ensuring both face and criterion validity, ways of calculating reliability. It was all very complex and it scared me. We learned that if you make a multiple-choice item, it needs to be piloted, tested using statistical measures, there needs to be a balancing item elsewhere in the test to check the respondent’s performance. There were issues like test integrity and repeated measures design. It was too much. When I got into teaching it was a relief to just deal with the individuals in the class and ignore all the crazy intensity of test construction.

Steve’s view of the complexity psychometric testing as analogous to language assessment is probably unreasonable, but this turn embodies the view we hold of the complexity of assessment which forms much of our initial discussion on our LAL.

A final point on this first reflection evidently shows our willingness to share openly on our experience and identify areas of weakness in our professional lives, regardless of our geographic distance and the remote platform. We wondered, in fact, if the remote platform itself served to increase the need to be careful in sharing our experiences, and thereby facilitate greater reflection. As we moved forward with our reflections together, we found our discussions became even more open and wide-ranging.

4.2. Reflection 2: Our LAL in Covid-19

Talk of the impact of Covid-19 on assessment forms the focus of our second reflection but is found across the data. In this section, the extracts selection is not a sequential dialogue. Instead, we have selected turns best representative of the three themes which emerged from our perspectives of the Covid pandemic on our LAL.

Wenwen (Reflection 2 turn 1): At the beginning stage of Covid-19, I moved back to China from Thailand and worked from home for a Thai university [which] implemented its distance learning contingency plans connecting students and teachers through online platforms and tools, such as Zoom, Google classroom, Microsoft Team, etc. During the lockdown period, I felt forced to quickly adapt myself to the “new reality” for the physically unreachable students and our virtual future. I was questioning myself how I would ensure that students
could be engaged in learning and how I could assess their learning. Before Covid-19, the courses I taught adopted a mixed means of assessment, that is, combining summative evaluation which informs teachers of students’ success or failure in their learning process based on a numeric scale and formative assessment which helps teachers to collect detailed information that can be used to improve instruction and student learning while it’s happening. Considering the inconvenience and almost impossibility of conducting summative evaluation during the Covid situation, I was forced to, but also privileged to, design or adjust tasks and implement assessments based on my students’ levels and needs.

**Steve (Reflection 2 turn 28):** I had to do some snappy thinking to achieve what I had been doing before in my previous classes. Some things were easy to simply adapt. Tests, for example, could simply go onto a google doc. However, as soon as I started doing that, I realised that it wasn’t going to be so easy. Because I couldn’t monitor students as they did their assessment, I had to build into the task some way of knowing if they were cheating, or to find a way to make the task cheating proof. For my own feedback on student learning, I decided to focus much more on formative style assessments, which I hoped would increase student involvement in the class, give me opportunities to hear the students using English, and let me know what problems they were having with their language use. When this all came up, I didn’t really know how to do it, but I got a great idea from Russell Standard’s videos on using e-portfolios. I watched this and then started using these ideas to get a portfolio system going for the students - as much as I could anyway, considering that they are also learning the technology while I am.

**Kamal (Reflection 2 turn 30):** In my case, the faculty was offered an optional training course for online assessment (by COURSEERA in collaboration with the Commonwealth of Learning) through Athabasca University Canada. In this course, there were many handy tips regarding using technology for online assessment but it really did not work mainly because:

- The online gadgets were not effective. Mostly there were connection issues and the internet was low.
- The number of students in each class was so large that it was not possible for a teacher to complete online assessment and evaluation for all students and all classes.
- Since I took the training along with most of my colleagues, we got our assessment skills improved in many ways but particularly in terms of:
  - Our use of technology i.e., tools for online assessment, e-portfolio, Zoom, Screencastomatic, Google Drive etc., and,
  - Our understanding of the philosophy of online learning i.e., it is formative, student-centric and collaborative.
- Despite our increase in assessment skills, in practice it was not applied appropriately because of the connection issues and large number of students.

True to our goal, our second reflection highlights our responses to the massive disruption brought about by Covid-19 and the subsequent need for experimentation and changes. We identify the failure of established procedures and approaches in dealing with assessment under the Covid conditions, and a resulting imperative to search for new, relevant solutions. Wenwen and Steve compare pre- and during-Covid techniques, and outline ways to achieve what was done before the lockdown with available alternatives. This search for a response to Covid is more systematic in Pakistan, where training courses in online assessment are made available to Kamal and his colleagues. Kamal’s reflection is notable in that the responses to Covid-19 formed part of a faculty-wide effort (‘… along with my colleagues’) rather than an individual endeavor, as it is with Wenwen and Steve. The shared response to the challenges in Kamal’s context is evident in the first reflection with his use of inclusive plural pronouns, but in this reflection the team’s involvement in dealing with the difficulties is made explicit.

Leading from the failure of established systems and the subsequent search for alternatives, there are indications of greater individual agency. In the first reflection, we see expressions indicating a lack of agency, for instance in the way we reported on our training, ‘I have never been trained’ (Wenwen) and ‘we were not properly given chances’ (Kamal). In the test-driven contexts of China and Pakistan, where authorities traditionally and closely manage assessment, teachers may feel they have little to contribute to their own LAL. However, with the onset of the pandemic, authorities had difficulty keeping pace with the demands for changes, thereby giving teachers the ‘privilege to design or adjust tasks’ (Wenwen). This increased sense of agency is easily identified in Wenwen’s extract (I was questioning myself), but is somewhat lost in the shared responsibility given by Kamal’s community response. In Pakistan, the increased control over assessment-related tasks was sudden and overwhelming, butting teachers under great pressure:

**Kamal (Reflection 1 turn 34):** During the Covid-19 situation our challenges now also include creating technology-enabled assessment and evaluation (alternative assessment). We’re not trained for technology-enabled assessment so teachers are facing a lot of issues and they are not only confused but have also lost confident of their skills and strategies.

Another consequence of the move to remote assessment was a push for greater digital literacy. Each of us speaks about the need to update both our assessment literacy and our burgeoning understanding of ‘online gadgets’. Steve found help in YouTubers (Russell Stannard) on a new approach (e-portfolios), Wenwen refers to the ‘platforms and tools’ used in her various courses, and Kamal’s training course provides ‘handy tips regarding using technology’. These are all aspects of our profession which we would not have accessed without the demands made by the shift online. We interpret this push for greater literacy as one of the positive outcomes of the move to online teaching and assessment.

The benefits of increased agency and push for greater digital literacy brought about by the move online required was not immediately obvious to us during our reflections, and in the analysis of the data they were marginalized by the far more noisy evidence in our discussions on the various constraints we faced: ‘the impossibility of conducting summative evaluation’ (Wenwen), ‘I couldn’t monitor
students’ (Steve), and ‘connection issues’ (Kamal). Understandably, in our search for solutions, the challenges facing us as assessment practitioners represented a serious source of frustration and formed the focus of much of the reflection. Our preoccupation with these issues continued into the third reflection and will be discussed in more detail in the next section.

One final point worth making about this second reflection is the unexpected depth of our assessment literacy that forms part of our discussions on these constraints. In our first reflection, we each noted our reservations regarding the strength of our LAL. On analysis of the data, however, we were surprised by the confidence with which we described the various kinds of assessment we needed to undertake. Admittedly not highly technical, our discussions do show sufficient knowledge and background to identify assessment problems and engage in evaluating the possible solutions available.

4.3. Reflection 3: Post Covid-19

The third reflection aimed to be a reflection for actions following our Covid-19 experience, and therefore to focus on our future assessment practices. However, perhaps because we were still too immersed in the Covid-19 situation, our reflections circled around many of the issues that we had already visited in our previous reflections. What dominated was not an exploration of future practice, but an impassioned exploration of various constraints on our assessment practice: student cheating in online tests and the difficulty of testing very large groups each elicited particular invective. By this stage in our interactions, our meetings had become a rich arena for sharing.

At the time of this third reflection, tensions around assessment had become a national issue in Pakistan, where students were rioting in response to decisions about exam writing venues (see www.dawn.com/news/1603740 for an example). For Kamal, this had immediate relevance. It thrust into the foreground many of the issues for educators that had formed part of our reflections. In our third Zoom meeting, Kamal attempted to present the complexities of the situation. The following turn from Kamal’s explanation is reproduced in full.

Kamal (Zoom meeting turn 45): Now in my country students are really protesting, would you believe this, they are demanding an online exam, because the universities decided to make the examination face to face and the students are not ready for this because there is a possibility of cheating in the online exam system. You know the online examination philosophy is a totally different thing. I had one training in my university in the first semester of Covid-19 by a university in Canada. They were training us for the new exam system which was a technological examination, and they were teaching us those principles of the online exam system, that is a student inclusive exam, the students are supposed to be part of the examination system. And you as a teacher are immediately required to shift to the new system, right? But students were not really providing the feedback to us practically so we didn’t have any other option but to go for an exam. In the previous philosophy there was a scoring system which is a must, we should submit at the end of the exam results sheet to the controller of examinations. Right. So that was the real problem with us teachers. When the students got to the new system, they did very well in the online exam system because they weren’t afraid to do anything. The teachers were unable to observe students individually, you know, because of the greater strength of the students, one teacher was supposed to be managing more than 60–70 students. So it wasn’t really in the control of the teachers. And as a result, most of the students were attempting their question papers with the help of other students, colleagues, peers, and with the help of books, and with the help of lots and lots of websites.

This analysis does not aim to unpack the complexities facing the Pakistani authorities. It only explores Kamal’s emic response. His description presents teachers as having to deal with three challenges: the theoretical solution to examinations (embodied in the ‘student-inclusive’ examination program) which teachers were ‘immediately required’ to adopt; student ennui (their lack of feedback to the new system); and bureaucratically predetermined requirements for scores (to the ‘controller of examinations’). Frustrated, subject to contradictory expectations, and hampered by the need to supervise very large numbers of students, the teachers are stripped of the power normally given to them through the high-stakes examination processes. This disruption to teachers’ authority occurs at two levels. Firstly, the students’ move to take control of their examinations challenges teachers’ role in the examination process. Secondly, the possibility of student cheating, resulting from poor examination supervision, undermines the value scores play in the educational system. The shift to online examinations, then, represents a major disruption to traditional power relations within the educational hierarchy.

In the context of this study, Kamal’s experience aligns with two of the themes that emerged in the earlier reflections. First, the role of training and the theory-practice disjunct. Kamal describes systematic and proactive training of teachers for ‘the new exam system’. Evidently based on solid theoretical principles, the program appears to have succeeded in preparing teachers for a major paradigm shift. Implementation of the training, however, is obstructed by the students who fail to contribute as expected, and from the administration which continues to require the ‘exam results sheet’. As we found in reflection 1, formal training is highly valued, but the implementation of theoretically sound assessment strategies is complicated in practice, creating a tension between theory and practice.

The second theme from Kamal’s narrative relates to teacher agency in the context of online assessment. As discussed, moving online facilitated greater independence from central bureaucracies in the rush to find ways to achieve online what had been the norm in class. In the conflict created over online assessment, much of the frustration expressed by Kamal centers around a sense of helplessness as teachers find their growing sense of agency undermined by opposing demands. At the individual level we saw how Steve and Wenwen were able to take control over their personal teaching contexts, but at the national level Kamal shows how hierarchically determined policy choices continues to impact the teacher at the classroom (or computer screen) level.

Much of our third reflection focused on the constraints brought about by the Covid lockdowns and the resulting challenges with online assessment. Kamal’s extract refers to two of these: cheating and large class sizes. Where Steve works with small groups and can successfully implement formative assessments which can mitigate cheating (such as e-portfolios in Reflection 2), Wenwen and Kamal
rely on mass summative assessments to achieve their institutionally required scoring. As we brought our final reflection to an end, these constraints remained unsolved and our goal of reflecting-for-action on our LAL in a post-Covid world incomplete.

5. Discussion

The use of duoethnography as a method for exploring our LAL allowed each of us to explore our own perspectives through a process of reflective dialogue. Our method of compiling a shared reflective document and following up with face-to-face discussions served two functions. First, we were able to reflect on our practices and experiences in our own time but with reference to the reflections of the others. We could therefore position ourselves and our experience in terms of shared experiences. This happened largely through comparing and contrasting our own lived experience with that of the others. Second, the face-to-face meeting allowed for much greater dialogic interaction during which we could build on the similarities in our experiences, but also challenge our narrative through the active participation of our peer listeners, leading to what Rose and Montakantiwong (2018) refer to as joint reflective practice. It must be noted that this focus on our individual lived experience is necessarily subjective and unrepresentative of all professionals in our context. In fact, the findings indicate that instead of their being a single experience, there were shifts in our understanding and subjective experience of LAL even as the research unfolded within the dialogue itself.

This shift in personal understanding is most clearly seen in the changing views of our own LAL as the project progressed. In reflection 1 there are explicit expressions of our own insecurity and lack of confidence in our individual LAL despite our teaching experience. However, the data show a mismatch between what we perceived as poor LAL knowledge and skills and the rather vigorous discussions around language assessment practice. As we shared our experiences and discussed our solutions, a growing sense of our own LAL competence emerges through the reinterpretation of ourselves in our narratives, and through the dialogic exchange (Babaii & Asadnia, 2019; Xu & Brown, 2016). This growing confidence plays out in our willingness to confront and explore sensitive questions relating to assessment practice, supporting Scarino’s (2017) argument that teachers each need to develop a personal framework for their LAL. Our data highlights, then, the value of reflective dialogue as crucial to this shift from insecurity to a willingness to participate in and confront questions of language assessment (Kato & Mynard, 2015). That this reflection was so successful in this regard may highlight the value of dialogue between colleagues in countries and contexts who can serve as disinterested listening partners (Blau & Fingerman, 2009), especially teachers and practitioners working alone or in combative environments (Richards, Hemphill, & Templin, 2018).

A recurring theme in the data is the role and value of training in our LAL. Our references to our training were made with a sense of pride and ownership, and the lessons learned appear to have been valued. However, we express awareness of shortcomings in the training and identify what we experience as a mismatch between the training, which represents formal theoretical values, and the raw demands of the practice of language assessment. We saw in Reflection 3, for example, that valuable training outcomes were obstructed by the incompatible objectives of the student body and the administration. These findings, we feel, speak in favor of the benefits of training in developing teachers’ LAL (Xu & Brown, 2016), but that its value would be enhanced if it were better matched the reality of the teachers’ assessment realities and account for the specific constraints teachers are likely to experience in their LAL practice.

This study took place a year into the pandemic when it was becoming clear that the lockdowns and shifts to online education were unlikely to be a temporary adjustment. Covid-19 presents a ‘critical incident’ for us as language assessment practitioners. The disruption caused by the pandemic and its resultant shutdown of schools affected almost every aspect of teaching practice. Our reflections on the effect of the disruption to our assessment practices in our reflections may serve as a catalyst for transforming the event itself into a site for rethinking ourselves as assessment practitioners and revisiting our language assessment experience:

**Steve (Reflection 3, turn 5):** Under normal circumstances, these are difficult questions to answer in a language classroom, but with online teaching it becomes almost impossible - especially if students’ cameras are off, or the class is very large, or if they are watching videos at home and only posting responses to the tasks. In this new paradigm, the concept of feedback on students’ learning through tests needs to be completely revisited. How can we do that when the scores we have to give still fit with a paradigm that has now been superseded?

6. Conclusion

This study documents how reflection, as a critical dimension of teachers’ professional development (Babaii & Asadnia, 2019), could assist English language teachers of different sociocultural contexts examine their understanding of LAL, reassess language assessment practices, and thereby enrich their repertoires of language assessment practice. We have shown that through the stages of our reflective dialogues, we became more confident in our LAL, took greater control of our assessment dilemmas, and grew more aware of embracing a perspective of language assessment in which alternative assessments could be tailored to our pedagogical needs. Although this study is mainly situated in the unique sociocultural contexts of three Asian countries, it addresses the global issue reflection as a key to transformational professional development. It is hoped that this study can motivate teachers across cultures to share their assessment practices with colleagues home and abroad on designing meaningful assessment tasks and developing effective alternative assessments. Future research is needed to explore how to address language teachers’ reported dissatisfaction with their own assessment literacy (Abrar-ul-Hassan & Douglas, 2020) and how to bridge the theory-practice disjunct of language assessment. It is recommended that LAL stakeholders, especially front-line teachers, be given training opportunities to develop their LAL through interacting with other professionals in an environment of collaborative reflection.
Author statement

**Stephen Louw**: Conceptualization; Methodology; Data curation; Data analysis; Resources; Validation; Writing-original draft; Writing-review & editing.

**Muhammad Kamal Khan**: Conceptualization; Data analysis; Investigation; Validation; Resources; Review & editing.

References

Abrahamson, D., & Chase, K. (2015). Interfacing practices: Domain theory emerges via collaborative reflection. *Reflective Practice, 16*(3), 372–389. https://doi.org/10.1080/14623943.2015.1052384

Abrah-ul-Hassan, S., & Douglas, D. (2020). Assessment and good language teachers. In G. Carol, & T. Zia (Eds.), *Lessons from good language teachers* (pp. 107–120). Cambridge: Cambridge University Press.

Ashraf, A., & Zaki, S. (2019). Language assessment in higher education: Challenges and consequences. *Pakistan Journal of Education, 36*(3), 163–185. https://doi.org/10.3909/pje.v36i3.437

Babaii, E., & Asadnia, F. (2019). A long walk to language assessment literacy: EFL teachers’ reflection on language assessment research and practice. *Reflective Practice, 6*(1), 1–16. https://doi.org/10.1080/14623943.2019.1688779

Banegash, D. L., & Gerlich, B. (2021). Critical language teacher education: A duaoethnography of teacher educators’ identities and agency. *System, 98*, 102474. https://doi.org/10.1016/j.system.2021.102474

Blau, M., & Fingerman, K. L. (2009). *Consequential strangers: The power of people who don’t seem to matter... but really do*. WW Norton & Co.

Blommaert, J., & Dong, J. (2020). *Ethnographic fieldwork: A beginner’s guide* (2nd ed.). Bristol: Multilingual Matters. https://doi.org/10.21832/9781788927147

Bozkurt, A., & Sharma, R. C. (2020). Emergency remote teaching in a time of global crisis due to Coronavirus pandemic. *Asian Journal of Distance Education, 15*(1), 1–6. https://doi.org/10.5281/zenodo.3779083

Brookfield, S. D. (1990). *Using critical incidents to explore learners’ assumptions*. In J. Mezirow (Ed.), *Fostering critical reflection in adulthood*. San Francisco: Jossey Bass.

Burhan-Horasanlı, E., & Ortaçtepe, D. (2016). Reflective practice-oriented online discussions: A study on EFL teachers’ reflection on language assessment research and practice. *Babaii, E., & Asadnia, F. (2019)*.

Coffman, J. (2005). *Private higher education in Pakistan: The need for order*. In *Reflective dialogue: Advising in language learning. Routledge*.

Coomeb, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia, 10*(3), 2–16. https://doi.org/10.1177/1462394319890970

Farrell, T. S. C. (2008). Critical incidents in ELT initial teacher training. *ELT Journal, 62*(3), 3–10. https://doi.org/10.1093/elt/ccm072.1

Farrell, T. S. C. (2012). Reflecting on reflective practice: (Re)visiting Dewey and Schon. *TESOL Journal, 3*(1), 7–16. https://doi.org/10.1002/tesj.10

Farrell, T. S. C. (2013). Critical incident analysis through narrative reflective practice: A case study. *Iranian Journal of Language Teaching Research, 1*, 79–89. Retrieved from https://ijltr.urmia.ac.ir/article_20454.html

Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *How, 26*(1), 35–61. https://doi.org/10.19183/how.181

Herrera, M. L., & Macias, V. D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal, 17*(2), 302–312. https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09

Hidri, S. (Ed.). (2021). *Perspectives on language assessment literacy: Challenges for improved student learning* (1st ed.). Routledge. https://doi.org/10.4324/9781000316083

Huang, Z., Wang, T., & Li, X. (2015). The political dynamics of educational changes in China. *Language Testing, 25*(3), 385–402. https://doi.org/10.1177/0265532209011064

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385–402. https://doi.org/10.1177/0265532209011064

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing, 27*(4), 555–584. https://doi.org/10.1177/0265532210363431

Kato, S., & Mynard, J. (2015). *Reflective dialogue: Advising in language learning*. Routledge.

Lantolf, J. P., & Johnson, K. E. (2007). *Extending Firth and Wagner’s assumptions*. In J. Mezirow (Ed.), *Fostering critical reflection in adulthood*. San Francisco: Jossey Bass.

Lowrie, J. R., & Kiczkowiak, M. (2016). Native-speakerism and the complexity of personal experience: A duaoethnographic study. *Cogent Education, 3*(1). https://doi.org/10.1080/2331186X.2016.1264178.2019.1672173

van Manen, M. (1991). *The tact of teaching: The meaning of pedagogical thoughtfulness*. New York: State University of New York Press.

Murray, D. M. (1982). Teaching the other self: The writer’s first reader. *College Composition & Communication, 33*(2), 140–147. https://doi.org/10.2307/357621

Norris, J., & Sawyer, R. D. (2012a). Towards a dialogic methodology. In J. Norris, R. D. Sawyer, & D. E. Lund (Eds.), *Duoethnography: Dialogic methods for social, health and educational research* (pp. 9–40). Walnut Creek, CA: Left Coast Press Inc.

Norris, J., & Sawyer, R. D. (2012b). Why duaoethnography: Thoughts on the dialogues. In J. Norris, R. D. Sawyer, & D. E. Lund (Eds.), *Duoethnography: Dialogic methods for social, health and educational research* (pp. 289–305). Walnut Creek: Left Coast.

Richards, J. C., & Farrell, T. S. C. (2016). *Professional development for language teachers*. Cambridge: CUP.

Richards, K. A. R., Hemphill, M. A., & Tempkin, T. J. (2018). Personal and contextual factors related to teachers’ experience with stress and burnout. *Teachers and Teaching, 24*(7), 767–787. https://doi.org/10.1080/13540620.2018.1476337

Rose, H., & Montakantiwong, A. (2018). A tale of two teachers: A duaoethnography of the realistic and idealistic successes and failures of teaching English as an international language. *RELJ Journal, 49*, 88–101. https://doi.org/10.1077/0033688217746206

Sawyer, R. D., & Norris, J. (2013). *Duoethnography*. New York, NY: Oxford University Press.

Sebok, S., & Woods, J. C. (2016). Using duoethnography to cultivate an understanding of professionalism: Developing insights into theory, practice, and self through interdisciplinary conversations. In R. D. Sawyer, & J. Norris (Eds.), *Interdisciplinary reflective practice through duoethnography. Examples for educators* (pp. 163–181). New York: Palgrave.

Stiggins, R. J. (1991). *Assessment literacy*. *Phi Delta Kappan, 72*, 534–539.
Stiggins, R. J. (1995). Assessment literacy for the 21st century. Phi Delta Kappan, 77(3), 238–245. https://doi.org/10.2307/20405538
Taylor, L. (2009). Developing assessment literacy. Annual Review of Applied Linguistics, 29, 21–36. https://doi.org/10.1017/S026719050900035
Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. Language Testing, 30(3), 403–412. https://doi.org/10.1177/0265532213480338
Tripp, D. (1993). Critical incidents in teaching: Developing professional judgement. London: Routledge.
Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. Papers in Language Testing and Assessment, 6(1), 41–63.
Watson Todd, R. (2019). How is English assessed at Thai schools? THAITESOL Journal, 32(1), 1–15.
Xu, Y. (2019). English language teacher assessment literacy in practice. In X. Gao (Ed.), Second handbook of English language teaching (pp. 517–539). Springer.
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
Yu, W. M. (2018). Critical incidents as a reflective tool for professional development: An experience with in-service teachers. Reflective Practice, 19(6), 763–776. https://doi.org/10.1080/14623943.2018.1539652
Zhang, C., Yan, X., & Wang, J. (2021). EFL teachers’ online assessment practices during the COVID-19 pandemic: Changes and mediating factors. The Asia-Pacific Education Researcher, 1–9. https://doi.org/10.1007/s40299-021-00589-3

Web links
Russell, S. https://www.teachertrainingvideos.com/assessing_online.
UNESCO. (2020). COVID-19 impact on education. UNESCO. Retrieved from https://en.unesco.org/covid19/educationresponse. (Accessed 15 February 2020).