rEFLections
Vol 30, No 2, May - August 2023

Digitally-mediated Language Assessment Practice (D-LAP): Qualitative Case Studies of Four Thai EFL University Lecturers

PARIWAT IMSA-ARD*
Language Institute, Thammasat University, Thailand
SUPONG TANGKIENGSIRISIN
Language Institute, Thammasat University, Thailand
Corresponding author email: pimsaard@gmail.com

Article information
Article history:
Received: 3 Mar 2023
Accepted: 3 May 2023
Available online: 11 Aug 2023

Keywords:
Language assessment literacy
Formative assessment
Digital assessment
University lecturers
Higher education

Abstract

The escalating incorporation of digital pedagogical technology in higher education, particularly in the post-pandemic period, posits a potential evolution in the paradigm of language instruction and assessment. The focal point of this research is to uncover the digitally-mediated language assessment practices (D-LAP) employed by Thai EFL lecturers in the university context. To fulfill this purpose, four EFL lecturers from Thai universities were purposefully selected to partake in classroom observations and subsequent stimulated recall sessions. The primary objective of these classroom observations was to shed light on the participants' assessment practices. Furthermore, the subsequent stimulated recall sessions were orchestrated to delve deeper into the participants' chosen assessment methods. The findings revealed that the majority of lecturers exhibited an intermediate degree of digital assessment literacy. This was primarily manifested in their propensity to implement assessment tasks within a digital platform and to collaboratively formulate performance indicators with their students. Additionally, it was discerned that their assessment tasks were predominantly crafted based on those presented in their sourcebooks in order to uphold the instructors' consistency in all course sections. The primary intention of these assessment tasks was to assess students' learning and diagnose their knowledge, employing a diverse assortment of assessment methods. This research endeavors to offer meaningful contributions towards the augmentation of professional development programs centered around digitally-mediated language assessment.

INTRODUCTION

One of the most important duties of English language instructors is to assess their students' performance in the classroom. The work of Black and Wiliam (1998), which emphasises the significance of assessment in fostering learning, has placed classroom emphasis on language assessment (Fulcher, 2012). As a corollary, language instructors are anticipated to be knowledgeable and experienced in a variety of assessment components, including assessment literacy, in order to achieve the desired assessment goals. In light of this issue, instructors' language assessment

literacy is crucial for enhancing student learning (Fulcher, 2012). Numerous empirical studies (e.g., Fulcher, 2012; Lam, 2015; Watmani et al., 2020) demonstrate that language instructors lack proper testing and assessment knowledge and skills for implementing effective and successful assessment practice. Similarly, existing research (e.g., DeLuca & Johnson, 2017; Yamtim & Wongwanich, 2014) has shown that EFL instructors lack assessment knowledge, especially regarding the application of such knowledge into practice. This seems to confirm Vogt and Tsagari's (2014) assertion that some instructors had inadequate knowledge of language assessment for practical application. Moreover, the strong influence of exam-oriented culture in Asian contexts, particularly in Thailand (Imsa-ard, 2020; Kaur et al., 2016), supports the function of testing and assessment in the way that teacher practices emphasise just test-taking strategies while disregarding what is not on the test. These procedures entail the necessary competence in language testing and assessment, resulting in a gap between theory and practice. Many language instructors had little or no formal training in assessment (Berry et al., 2019), and their assessment practice may have been influenced by their prior testing experiences, causing them to test in the way they had been assessed (Smith et al., 2014; Vogt & Tsagari, 2014).

Moreover, with the emergence of 21st century skills, today's students and teachers must have a better knowledge of digital technology as it grows more prevalent (computers, electronic white boards, GPS, etc.). Despite the advancement of Information and Communication Technologies (ICT), there had not been a significant shift in language assessment practices, at least not until the COVID-19 began its worldwide assault. It is probable that such a dramatic shift in assessment practices reveals how difficult it has become to assess students' performance on online platforms. Under these conditions, language instructors must embrace digital literacy as well as alternative assessment modalities that are suitable for the new reality. Despite the fact that a pandemic is transient and the education system's evolution after a crisis is unpredictable, the COVID-19 pandemic triggered a digital revolution in academic and higher education that has continued to this day and heralded the beginning of the post-COVID-19 era (Alhat, 2020). After COVID-19, the teaching and learning environment, curriculum, and teaching practices will undergo a period of transition as a result of the new challenges it will bring about. This transition will culminate in the introduction and adoption of digital tools to support teaching and learning, despite the possibility that classrooms will return to face-to-face settings. Therefore, it is essential to investigate lecturers' digitally-mediated language assessment practice (D-LAP) in order to ensure that they are appropriately prepared for the post-COVID-19 era. Surprisingly, many recent research studies (e.g., Jan-nesar et al., 2020; Sultana, 2019; Xu & Brown, 2017) have found that assessment literacy among language instructors was still undeveloped, necessitating more studies on language teacher assessment literacy in many contexts. Moreover, the circumstances caused by COVID-19 have raised significant concerns regarding assessment among students, parents, and teachers (Duraku & Hoxa, 2020; Maaoui et al., 2023). In line with such concerns, Astiandani and Anam (2021) demonstrates that EFL instructors need assistance with technological issues and socialising with digital devices for their assessment.

Also informed by the aforementioned discrepancy, such issues could possibly hamper the appropriate assessment practice aimed at promoting students' learning. It would be more

beneficial to investigate lecturers' digitally-mediated language assessment practice in order to establish a more comprehensive view of lecturers' digitally-mediated language assessment practice, which could broaden a comprehensive understanding and hopefully promote students' language learning. Despite the breadth of literature on assessment literacy on instructors, scant attention has been paid to investigating Thai EFL university lecturers' language assessment practices, particularly in terms of digitally-mediated language assessment. Therefore, this study attempted to investigate the Thai EFL university lecturers' digitally-mediated language assessment practice.

In light of its relevance, it is envisaged that this study will contribute to the EFL and language assessment research literature. Significantly, this research may assist instructors, educators, and trainers in providing effective support and encouragement for their assessment practice, especially in terms of digitally-mediated assessment, so as to improve students' long-term learning. In addition, the results will give illuminating information on teachers' perspectives and practices that may be used to enhance and support teacher professional development programs as necessary.

LITERATURE REVIEW

Digitally-mediated language assessment

Assessment quality is an essential aspect of instructors' assessment literacy since the primary goal of assessment literacy should be to improve assessment quality and maintain that assessments are utilised appropriately for their intended purposes. Given the connection between teaching and assessment, instructors need to acquire assessment principles in order to enhance their teaching and assessment abilities. Initially, the term assessment literacy emerged in Stiggins's work (1991). Stiggins (1995) emphasised the significance of assessment literacy to the quality of teaching and learning and urged for research in this area. Moreover, he emphasises that instructors are not just users of assessment data, but also need more sophisticated skills and knowledge to differentiate between high-quality and low-quality assessment practices. Since then, throughout the preceding three decades, a number of experts have focused on enhancing teachers' assessment literacy (Brookhart, 2011; Inbar-Lourie, 2008; Popham, 2009). Although there have been different definitions of assessment literacy, it is commonly agreed that assessment literacy requires teachers to be able to connect testing and assessment theory with the practice of testing and assessment. This includes their understanding and knowledge of the theoretical issues of assessment, as well as the appropriate application of assessment practices. Simply described, assessment literate instructors not only comprehend the theoretical issues of testing and assessment, but also understand how to put them into practice. Focusing on language teacher assessment literacy, Popham (2011) defines AL as "an individual understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions" (p. 267), which implies that teachers should be able to make inferences and draw conclusions about students based on their behavioural repertoire as well as their covert knowledge and skills. When drawing from general assessment literacy, Language Assessment Literacy (LAL) refers to stakeholders' familiarity with assessment practices,

as well as their application of this knowledge to classroom activities in general, and especially to concerns of language assessment (Taylor, 2009).

Furthermore, Brookhart (2011) presents a comprehensive and updated inventory of knowledge areas and skills that educators should acquire to proficiently and professionally undertake assessment-related tasks. The constituents of this list are grounded in the 1990 Standards, albeit significantly updated to reflect the evolving field of formative assessment. This compilation furnishes educators with a robust understanding of educational assessment principles that align with modern teacher assessment standards. These include: 1) Appreciating the dynamics of language learning and teaching, 2) Establishing and communicating clear learning objectives, 3) Conveying learning outcomes to students, alongside devising tasks that accurately embody these outcomes, 4) Utilizing various assessment methods and understanding their purposes, validity, fairness, and reliability, 5) Employing critical thinking skills, understanding cognitive levels, analyzing test items, and making informed instructional decisions based on student performance, 6) Articulating and disseminating constructive feedback, 7) Grading, performing item and quantitative analyses, and making informed decisions based on these assessments, 8) Administering and interpreting results from external assessments for informed decision-making, including interpreting scores from large-scale assessments and managing institutionally mandated assessments, 9) Communicating effectively with parents and colleagues about educational reforms, 10) Understanding the intricate relationship between assessment, student motivation, perception of authority, and self-management, as well as facilitating students in evaluating, planning, and monitoring their academic progress, and 11) Developing tests, maintaining confidentiality, and adhering to ethical assessment practices. This encompassing list of competencies is intended to support educators, professional development specialists, teacher trainers, and other individuals responsible for evaluating teachers' assessment knowledge and skills, bolstering their efforts in this critical undertaking.

In addition to possessing language assessment literacy, with the utilization of digital technology in education, it is required that language instructors should possess a so-called "digital assessment literacy" to be able to utilise a range of applications and technological systems to improve students while using a number of assessment strategies. Prior to 2012, the term "digital assessment literacy" had not been used in academic literature until Eyal (2012) invented the term, but it may have been suggested to refer to the role of the teacher as an assessor in a technology-rich environment. Eyal (2012) focuses on a particular aspect of assessment literacy: digital assessment literacy. On a three-tiered continuum, Eyal (2012) proposes the following abilities and skills that teachers must possess in order to achieve digital assessment literacy: basic, intermediate, and advanced. They were characterised as follows (Eyal, 2012, p. 45):

1) Digitally-mediated assessment literacy is a continuum, beginning with the most basic level, which involves the use of LMSs (Learning Management System) in conjunction with more traditional assessment processes, such as computerised tests.
2) In addition to traditional processes in assessment practice, the intermediate level involves the administration of examinations, tasks, and projects in a digital environment, with performance indicators created in collaboration with students.

3) At the advanced level, estimation approaches are constructed around constructivist-social learning and the growth of self-directed learning, while teachers must also understand how and when to delegate assessment processes to students.

In conclusion, the term "digitally-mediated language assessment" has been operationally defined in this research as a measure of assessing a one's language proficiency by utilizing technology and digital tools. However, there is an evident lack of research (e.g., Brown, 2013) that investigates practical skills of language instructors when it comes to utilizing technology for language testing in their classrooms, let alone investigating the use of computers in testing as part of a digitally-mediated assessment practice. Thus, further studies are needed to determine how each level of digital assessment literacy might be implemented in classroom practices.

Language assessment practices

Although Black and Wiliam (1998) demonstrate that classroom assessment is commonly understood to involve a variety of activities that teachers and students engage in to gather data that can be used diagnostically to improve teaching and learning, classroom assessment practices are widely believed to be multifaceted and multidimensional (Rodriguez, 2004). Prior to that, it is necessary to define the term "teachers' assessment practices". Stiggins et al. (2006) propose the following five characteristics of effective classroom assessment practices: "Clear purposes, Clear targets, Sound design, Effective communication, and Student involvement" (p. 27). To support Stiggins et al. (2006) regarding student involvement, Black and William (2009) also demonstrate that innovative language learning theories emphasise how students learn based on formative assessment information they receive and their interpretation, enabling them to make effective learning decisions. In light of this, student involvement in the use of assessment information necessitates an expansion of teachers' assessment knowledge and skills, as this will require teachers to be adept at articulating and sharing learning intentions and success criteria with their students, as well as providing meaningful opportunities for students to act on assessment information they receive (Cheng et al., 2004).

Subsequently, Brookhart (2011) proposes a set of assessment competences based on the distinct requirements that instructors must fulfil while doing assessment tasks: 1) Prior to completing assessment tasks, instructors must have both content and pedagogical knowledge, 2) Instructors should be able to describe the goals and objectives of their instruction and, by extension, their assessments, 3) Instructors should be able to communicate with students and colleagues on success criteria, especially when it comes to formative assessments, 4) Instructors must be aware with the purpose and use of various assessment techniques from either published coursebooks or teacher-developed techniques (Hosseini & Azarnoosh, 2014) and skilled in their application, 5) Instructors should be able to analyse classroom questions, test items, and performance assessment assignments to determine whether or not students have the essential knowledge and critical thinking skills, 6) Instructors must be able to offer students with effective and constructive feedback on their work, 7) Instructors must be capable of creating scoring systems that transform student performance into actionable data, 8) Instructors should be able to administer and interpret external assessments in order to make appropriate placement

decisions for students, 9) Instructors should be able to articulate their interpretations of assessment data and their justifications for making sound judgements, 10) Instructors should be able to support students in making sound educational decisions based on assessment information, and 11) Instructors should be aware of their legal and ethical obligations during all assessment procedures. Within this context, "assessment tasks" denote the explicit activities that students are mandated to undertake within the parameters of the assessment process. Such activities may encompass a spectrum of tasks, ranging from drafting an essay to delivering an oral presentation. Conversely, "assessment methods" pertain to the tactical strategies or methodological techniques leveraged by instructors to scrutinize and quantify student learning. It is crucial to understand that a singular assessment method may encapsulate multiple assessment tasks. For instance, within the context of a project-based assessment method, the embedded assessment tasks could potentially include a preliminary phase of topic research, followed by a written report, culminating in a presentation of the acquired findings.

In light of the profound digital disruption catalyzed by the COVID-19 pandemic, the integration of digital tools in instructional methodologies, pedagogical approaches, and assessment procedures has surged significantly. The transformational effect of technology on educational assessments can no longer be disregarded in the epoch of digitalization (Yu & Zhang, 2017). This evolving landscape has necessitated educators to acquire and develop novel competencies to navigate the challenges inherent to digital language assessment (Eyal, 2012; Taylor & Harding, n.d.). These challenges have been further amplified in the context of the COVID-19 pandemic. Notably, there remains an absence of terminologies specifically crafted to elucidate the practice of language assessment utilizing digital platforms. Nevertheless, Eyal (2012) proffered the concept of "digital assessment literacy," which alludes to the strategic employment of a myriad of applications and technological infrastructures to foster student progress. This concept encapsulates a broad spectrum of assessment strategies adapted to the digital milieu.

The degree of digital assessment literacy is construed as a continuum. At its nascent stage, it involves the incorporation of Learning Management Systems (LMS) into conventional assessment procedures, such as computerized examinations. Ascending along the literacy continuum, one encounters a more sophisticated degree of literacy. This entails, beyond traditional methods, the execution of tests, tasks, and projects within a digital framework. Performance indices for these tasks are collectively formulated with student input, underscoring a participatory approach. Culminating at the apex of this continuum is the integration of progressive evaluation methodologies premised on constructivist-social learning and the fostering of self-directed learning. This requires educators to possess a keen understanding of the opportune moments and effective strategies to delegate the assessment procedures to their students. Such a multi-faceted approach encapsulates the full range of digital assessment literacy, highlighting its essential role in the contemporary educational landscape.

Related studies

There have been a numerous studies investigating language assessment practices in several contexts. To begin with, Babaii and Asadnia (2019) examined five EFL teachers' reflections on their research-based language assessment knowledge (reflection-on-research), current language

assessment practice (reflection-on-action), and future language assessment plans as they progressed through their professional development (PD). Qualitative approaches included online discussions, reflective narratives, interviews, and scenario creation. Concerning a social networking platform for teachers to discuss the practical significance of new research articles to build community and professional development. Teachers were also asked to write three focused narratives on pair work, LAL, and diagnostic assessment in reflective narratives, which encouraged them to move beyond their pedagogical performance and emphasise their assessment practices. Finally, the researchers built the scenarios to assess how instructors use theories and reflect-for-action. Teachers were interviewed for LAL improvement suggestions in semi-structured interviews. Their findings demonstrated that instructors' comments on theories and language assessment practices were comparable, which may contradict previous literature suggesting discrepancies. Supervisors might also hold workshops and discussion panels to assist instructors improve their LAL and provide constructive assessment feedback. This study suggests that teachers' reflective journals can help them improve their classroom-based assessment practices and that familiarising them with digital language assessment technologies like mobile- and computer-assisted language assessment can help them improve their LAL.

While attempts have been made to investigate teachers' language assessment practice, language teachers have been confronted with several challenges, including the forced transition to online teaching (i.e., emergency remote teaching) in response to the COVID-19 pandemic. As a result, it is critical to review studies on online/digital language assessment practices. To begin with, Nam (2021) examined 124 graduate teaching assistants' (GTA) online assessment practices. The findings showed that online teachers used open-book examinations, video recording, and written and audio-recorded feedback. However, teachers reported that they wanted more hands-on language teaching and assessment training. In addition, Chung and Choi (2021) examined an English language program in South Korea to discover how the abrupt shift to online language instruction has affected 32 full-time language teachers' teaching and assessment practices. They conducted a speaking-intensive course to help students enhance their English communication skills and prepare them to utilise English long-term. Teachers assigned multimodal projects to generate sustainable assessments in which students may actively use target language forms and structures. In addition, Al Bahlani (2019) utilised mixed methods to investigate EFL teacher assessment literacy in Omani colleges. This study assessed EFL teachers' assessment literacy by examining their perceptions of their assessment competence and practice, their performance on an assessment knowledge test, their performance on a language test's assessment evaluation task, and their actual assessment practices in the language classroom. Self-assessment surveys, a language assessment knowledge test, an assessment evaluation task, classroom observations with a focus on teacher-created assessment activities, and teacher interviews were used to measure teachers' digital assessment literacy. The study's context enabled the use of cutting-edge technology in language classrooms, but digital assessment literacy was poor.

According to a review of recent literature, a number of studies on language assessment literacy have been conducted from the perspective of instructors and at various teaching levels and in a variety of circumstances. Studies on lecturers' digitally-mediated assessment practices

(D-LAP) have received scant attention, and none have focused on Thai EFL university lecturers in Thailand. This creates a clear research gap that will be addressed by this study. The purpose of this study is to explore the digitally-mediated assessment practice of Thai EFL university lecturers in order to fill in gaps in the literature and to acquire insights into instructors' assessment practices in the hopes of further enhancing students' language learning in the digital age. In order to fill in the gaps, below is the research question: How do Thai EFL lecturers employ digitally-mediated assessment practices in their classrooms?

RESEARCH METHOD

Research design

In order to gain insight into Thai EFL university lecturers' digitally-mediated language assessment practices, the qualitative case study approach through classroom observations with stimulated recalls was espoused in this study. This is because "case studies are the preferred strategy when 'how' and 'why' questions are being posed, when the investigator has little control over events, and when the focus is on a contemporary phenomenon within some real-life contexts" (Yin, 1994, p. 1). Although this case study approach can be criticised for its lack of generalizability from one case to another due to the different contexts (Creswell, 2013; Patton, 2002), it enabled researchers to zoom in on the real-world context of language classrooms in order to comprehend and transfer how Thai EFL university lecturers assess their students to their own contexts. Since the concept of 'transferability' does not imply broad claims, the purpose of this case study research is to allow readers to make connections between the findings of this study and their own experiences and contexts.

Participants

A sample of four EFL university lecturers was selected using purposeful sampling. The four participants were chosen by using purposeful random sampling, which placed an emphasis on choosing "information-rich cases" (Patton, 2002, p. 46), providing 'in-depth' understanding and insights into the finding instead of empirical generalisation (Liamputtong, 2009). Qualitative research often makes use of purposeful sampling to choose and select samples that include a substantial amount of information on the topic of interest (Palinkas et al., 2015). More specifically, purposeful sampling, even with tiny samples, can significantly increase the credibility of research studies (Patton, 2002). The selection criterion for recruiting the participants in this study was based on the assessment knowledge test scores from the Phase 1 study (Imsa-ard, 2023). Significantly, the four participants participated on a voluntary basis. Thus, the participants in this study included two lecturers whose test scores were higher than a mean score, and two lecturers whose test scores were lower than a mean score.

The following table presents an overview of each participant, including their personal information, work experience and their context. To assure their confidentiality and anonymity, the participants' names were pseudonyms. The participants included two male lecturers (50%) and two female lecturers (50%). Their teaching experience ranged from 5 years to 13 years. Three of them (75%) hold master's degrees, while only one participant (25%) holds a doctorate degree.

Table 1

Demographic information of participants

Personal information				Experience	University context	
Name	**Gender**	**Age**	**Degree**	**Years of teaching in a tertiary context**	**Course observed**	**No. of students in the observed class**
Suda	F	35	M.A. in English as an International Language B.A. in English	10	English for Oral Communication	20
Chain	M	34	M.A. in English Language Teaching B.Ed. in English	5	English in the Digital Era	35
Mali	F	41	Ph.D. in Applied Linguistics	13	English for University Life	38
Petch	M	44	M.Ed. in Teaching English as a Foreign Language	8	English Skill Development	28

In this section, all of the four participants were described as follows.

1) Suda was a university lecturer who graduated her master's degree in EIL (English as an International Language) from a university in Bangkok, Thailand. She was responsible for teaching English foundation courses and English for Specific Purposes courses for undergraduate students. She has been teaching for 10 years.
2) Chain was a university lecturer who holds his master's degree in English Language Teaching from a university in Bangkok, Thailand. He was responsible for teaching English foundation courses for undergraduate students. He has been teaching for five years. At his university, most students have digital gadgets such as iPads or laptops. At the stage of data collection, students mostly studied through an online platform. Thus, most students are familiar with using digital tools for their study.
3) Mali was an assistant professor who holds her PhD in applied linguistics from a university in the United States. She was responsible for teaching English foundation courses and English for Academic Purposes courses for both undergraduate and graduate students. She has been teaching for 13 years.
4) Petch was a university lecturer who holds his master's degree in Teaching English as a Foreign Language from a university in Thailand. He was responsible for teaching English foundation courses for undergraduate students at one public university. He has been teaching for eight years.

Context of the study

This study was conducted in the university context of Thailand, all of which are public universities

ranked in both the QS World University Rankings 2022 and the SCImago Institutions Rankings as the top 10 universities in Thailand. In general, there are two types of universities in Thailand, namely public and private universities. To elaborate on public universities, which is the focus of this study, there are two types of public universities, namely government universities and autonomous universities (Kantabutra & Tang, 2010). In response to the advent of the COVID-19 pandemic, numerous higher education institutions instituted policies pertaining to the incorporation of technology in language instruction and learning. Notwithstanding the improvements in the pandemic crisis, the perpetuated usage of technological platforms within classroom settings remains evident. Predominantly, universities have gravitated towards platforms such as Google Classroom and Microsoft Teams for pedagogical purposes, whereas Zoom has been predominantly harnessed for the facilitation of virtual meetings. Given the context, it is pertinent to delineate the specifics of the present investigation, which was conducted across four public universities. Consequently, detailed descriptions of these universities, their characteristics, and the specific deployment of technology within their educational environments are warranted to better elucidate the scope and findings of this study.

First of all, the university where Suda works is thought to be one of the most prestigious in Thailand, and it seems that students who obtain a good rank on their university entrance exams can be admitted. Moreover, this university is a comprehensive and research-intensive public and autonomous university. At this university, most students have digital gadgets such as iPads or tablets, and buildings are well-equipped with digital tools and internet connections.

Secondly, the university where Chain works is thought to be one of the top public universities in Bangkok. At this university, most students have digital gadgets such as iPads or laptops. At the stage of data collection, students mostly studied through an online platform. At his university, most students have digital gadgets such as iPads or laptops. At the stage of data collection, students mostly studied through an online platform. Thus, most students are familiar with using digital tools for their study.

Thirdly, the university where Mali works is thought to be a public research university in Bangkok, where it has expanded its subject areas to cover life sciences, science, engineering, social sciences, and humanities. At this university, most students have digital gadgets such as iPads or laptops, and buildings are well-equipped with digital tools and internet connections.

Finally, the university where Petch works is thought to be one of the leading Asian academic institutions in Thailand, with a world-class standard in producing graduates and creating bodies of knowledge. At this university, most students have digital gadgets such as iPads or laptops, and buildings are well-equipped with digital tools and internet connections. Thus, most students at the university where he works are familiar with using digital tools for their studies.

Research instruments

In this study, there are two major instruments to address the research questions, namely the classroom observation protocol (Appendix A) and the stimulated recall protocol (Appendix B).

Classroom observation

For the purposes of the questions in this study, the researcher carried out three lesson observations and took on the role of a non-participant observer, meaning the researcher observed every situation without ever intervening in what was being observed. To support the number of observations, van der Lans et al. (2016) demonstrate that three lesson observations, in particular, should be very beneficial for researchers who are interested in connecting classroom observations of teaching with other variables in their research, and three lesson observations are very applicable to school assessment practices. Observation literature (e.g., Bryman, 2008; Robson, 2002) has emphasised the significance of developing an observation schedule for recording the events of each observation. Thus, an observation schedule (*See* Appendix A) was developed that outlines a number of categories that are the focus of the observations based on concept-driven information from quantitative data about the types of assessment purposes and strategies utilised in the classroom by teachers. Also, they were developed based on the concepts of formative assessment theories (Black & Wiliam, 2009) and Brookhart's (2011) framework. However, it is crucial to acknowledge the potential pitfall of observer bias within observational research. To mitigate this risk, the implementation of a stimulated recall technique is often recommended. This technique can counterbalance the inherent bias by prompting the observer to recall and reflect on the observation with increased objectivity, enhancing the accuracy and reliability of the collected data.

Stimulated recalls

A video-stimulated recall (VSR), a method in which the researcher shows the participants a video of their own behaviour in order to prompt and boost their subsequent recall and interpretation of the event, was used in this study during the post-teaching interview (See Appendix B). Moreover, stimulated recalls help to uncover the reasoning and thinking processes behind their activities (Gass & Mackey, 2016). The use of stimulated recalls not only assists in the elaboration of teacher digitally-mediated assessment practices, but it also improves the validity and reliability of the data obtained. During the post-teaching interview, the stimulated recalls were used to clarify the classroom events and the interviews were audio-recorded to facilitate transcription and data analysis. Throughout the post-teaching interviews, the application of stimulated recall techniques served as a tactical instrument to explicate and interpret classroom dynamics. To bolster precision in transcription and augment thorough data analysis, the interview sessions were meticulously audio-recorded, thereby enhancing the accuracy and completeness of the collected data. The interviewees were presented with video excerpts serving as stimuli, aimed at aiding them in recollecting their digitally-mediated assessment practices in the classroom milieu. In the pursuit of responding to the third research question, and to uncover the salient themes discerned during observational sessions, the researcher probed the interviewees with a series of questions. These questions were targeted at unearthing key insights pertaining to (a) the sources of the assessment task, (b) the explicit objective of the assessment, (c) the degree of student engagement incorporated, (d) the provision of educator feedback post-assessment, and (e) the subsequent utilization of the assessment data, among other thematic points of interest.

Data collection

The data collection process began in August 2022. Qualitative data were collected from classroom observations with stimulated recalls during August-November 2022 (i.e., the first semester of Academic Year 2022). Following verbal consent from participants, an information sheet and consent form were created for each. Prior to initiating any data collection procedure, a short discussion with each participant was held to explain the steps of the method and the purpose of each step. They were advised that data collection involved four appointments, once every two weeks, with each meeting lasting no more than an hour. The meeting time and venue were determined by the participant's preference, availability, and convenience. Each participant was invited to schedule an observation and interview at a time and location that was convenient for them (Arksey & Knight, 1999). Accordingly, an interview appointment schedule was prepared and communicated with participants; they were notified that, two days prior to each session, a reminder was issued. To avoid any inconvenience, planned times were adjusted according to their availability.

As shown in Figure 1, each participant was observed three times during weeks 3–4, 6–7, and 9–10 of their undergraduate English language courses. All class instructions were meticulously recorded to support comprehensive data analysis, given their substantial duration of three hours each. These recordings were duly deleted once the data analysis phase was concluded. Importantly, all participants were adequately informed and consented to these recordings prior to their commencement. Following each observation in the three-hour session, stimulated recalls were used to facilitate the interpretation of classroom events. It should be noted that stimulated recalls should be done as soon as possible to maximise what the participants can remember and avoid memory decay (Gass & Mackey, 2016). Thus, in this study, stimulated recalls were adopted right after the class observation or within the week of observation.

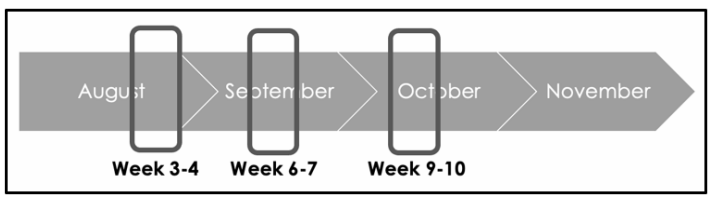

Figure 1 Data collection

Data analysis

In this study, content analysis was utilised to examine all qualitative data, including classroom observations and stimulated recalls with university lecturers. Firstly, the qualitative data gathered from classroom observations was analysed deductively using the observation schedule guided by the literature (Brookhart, 2011). This type of analysis was carried out to illustrate how Thai EFL university lecturers conduct assessment practice in their classes in aspects of 1) sources of assessment tasks, 2) purposes of assessment tasks, 3) assessment methods, 4) involvement between students and teacher, and 5) feedback in assessment practice. In addition, a part of the qualitative data was converted into quantitative data to provide descriptive statistics regarding assessment tasks used, methods, kind of feedback in the form of the

frequency of usage of teacher participants. Secondly, the analysis of stimulated recalls was then conducted by using content analysis. The data was inductively and deductively analysed. First, a deductive analysis was conducted to investigate Thai EFL university lecturers' digitally-mediated assessment practices in their classrooms in terms of sources of assessment tasks, purposes of assessment tasks, assessment methods, involvement between students and teacher, and feedback in assessment practice. To ensure the coding process is of high quality, the researcher should do inter-coder reliability analysis, which is regarded as a standard methodological requirement in content analysis (Krippendorff, 2011). In this study, the analysis was conducted by three coders, each of whom holds a minimum qualification of a master's degree in language education and possesses relevant research experience. Despite their high level of academic achievement, it is noteworthy that in order to minimize the potential for human error and subjectivity, they were provided with specific workshops centered on the study's objective.

Data obtained through stimulated recalls were integrated and triangulated with data collected through classroom observations, first with each individual lecturer and then collectively, to identify similarities and differences in their digitally-mediated language assessment practices.

FINDINGS AND DISCUSSION

To address the research question, the researchers looked at how Thai EFL university lecturers practised language assessment in their language classrooms. To be more specific, this qualitative case study approach was employed to look at the type of assessment purposes, strategies (methods and student involvement) that teachers practised in their language classes. The qualitative case study method was employed through structured classroom observations with stimulated recalls with four Thai EFL university lecturers. Albeit the analyses of classroom observations and interviews utilised an inductive method to identify emerging themes, they were primarily deductively analysed using categories derived from formative assessment theories (Black & Wiliam, 2009) and Brookhart's framework (2011): sources of assessment tasks, purposes of assessment tasks, and assessment methods, student involvement in the assessment task, and type of teacher feedback given on the task. In this part, the results will be presented and discussed according to the categories previously mentioned.

Sources of assessment tasks

To indicate the sources of each assessment task, stimulated recalls after each observation revealed that all lecturers developed their assessment tasks from a published coursebook used by their institutions. Suda indicated that all assessment tasks in her course were developed based on those in the coursebook. To elaborate, she also adapted some activities in the published coursebook to suit her students in the course as she stated:

> *"Since the course I am instructing is a foundation course for university undergraduates, I must rely on the published coursebook provided to each section. To make it appropriate for my students, I must adapt and modify the activities, that is to say the difficulty*

> *level, to their proficiency level while maintaining the course's objectives. This is to ensure that students could achieve the course outcomes."* (Suda)

This is also similar to what Petch and Chain mentioned:

> *"The bulk of my assessment tasks for this course have been based on the coursebook that every instructor uses. I avoid using my own assessment tasks since it may not be fair to my students if shared assessment tasks are used and modified for the exam. Utilising assessment tasks from published coursebooks would familiarise students with the assessment."* (Petch)

> *"Because every student in every section will have the same activities, I am compelled to use the assessment tasks from the published coursebook."* (Chain)

However, Mali added some teacher-developed tasks to supplement the course as follows:

> *"In fact, I rely mostly on the assessment tasks provided in the coursebook. However, I add certain teacher-developed and contextualised tasks to ensure that students have additional authentic practice opportunities."* (Mali)

From the aforementioned findings, it suggests that Thai EFL university lecturers mainly used an assessment task source from their coursebooks. This finding suggests that lecturers prefer to utilise the assessment tasks included in published coursebooks rather than create their own assessment tasks for their courses. This finding may be explained by the fact that the observed course is a required general education course for all first-year students; therefore, to ensure the reliability of instruction, lecturers must rely primarily on the assessment tasks provided in the coursebook, despite the fact that some lecturers also assigned supplementary tasks. It seems to corroborate Hosseini and Azarnoosh's (2014) findings demonstrating that the majority of the assessment tasks used by instructors were retrieved from published textbooks. Nonetheless, it seems to contradict Al Bahlani's (2019) results that EFL instructors used a range of assessment sources, particularly teacher-created assessment tasks were substantially more prevalent than coursebooks. This discrepancy is explained by the fact that participants in his research received a training session on how to better design assessment tasks that correspond to the level of their students and their intended goals, leading participants to develop their own assessment tasks. Significantly, it should be noted with caution that teacher-created assessment tasks may have issues with regard to task instruction clarity, validity, authenticity, and digital application.

Purposes of assessment tasks

In order to identify assessment purposes, each observed task was examined according to the assessment purposes given by lecturers during observations and stimulated recalls, because the majority of participants did not convey the purposes of the tasks overtly and directly during classroom observations.

Table 2

Purposes of assessment tasks

Assessment Purpose	No. of Tasks	Percentage
• Evaluating	48	38.71
• Diagnosing	35	28.23
• Reporting (feedback for students)	23	18.55
• Preparing for a coming class	10	8.06
• Evaluating previous learning	8	6.45

As demonstrated in Table 2, Thai EFL university lecturers mostly assessed and evaluated student learning through assessment tasks. In her stimulated recalls, for instance, Suda said that she utilised and adopted assessment tasks to "*mostly evaluate their learning in each lesson*" (Suda). This specific purpose was utilised in 48 (38.71%) of the 124 activities conducted by all participants. 35 tasks (28.23%) were administered to diagnose students' knowledge, while 23 tasks (18.55%) were administered to provide students with feedback. Early in the semester, for example, when Petch was asked about his assessment purpose, he said, "*I need to conduct some diagnostic assessment tasks to diagnose students' knowledge so that I can plan the lessons more effectively and students could know how much and what they know.*" (Petch). This finding is partially consistent with the findings of Cheng et al. (2004), whose participants similarly reported utilising assessment for more student-centred purposes. In a similar vein, Mali noted, "*Each assessment task could provide students with feedback so that they could recognize their own learning*" (Mali). Lesser proportions, 8.06 and 6.45 percent, represented the number of assignments that prepared students for upcoming classes and evaluated prior knowledge, respectively. A potential reason for such assessment purposes, e.g., evaluating and diagnosing students' knowledge, could be the lecturers' intentions to prepare their students to pass their summative exams at the end of the course, according to the stimulated recalls. This conclusion confirms Alkharusi's (2021) and Xie and Tan's (2019) arguments that the primary goal of classroom assessment is to evaluate and diagnose student performance in order to enhance their learning.

Assessment methods

The analyses used to establish the findings of this section were based on the observation schedule filled out by the researcher during observations and stimulated recalls for each observation. The assessment methods in this study were initially drawn from the scholarly literature (Black & Wiliam, 2009; Brown & Abeywickrama, 2018). The findings revealed that lecturers used a variety of assessment methods. Table 3 depicts the number and proportion of the various assessment methods used.

Table 3
Assessment methods

Assessment Methods	No. of Tasks	Percentage	Notes
• Fill-in-the-gap	47	37.90	Mostly, this method was adopted in listening and vocabulary tasks.
• Short answers	39	31.45	Sometimes, this method was employed with oral discussion to stimulate students' thinking.
• Oral discussion	26	20.97	Mostly, this method was utilised in group work and accompanied by short answers.
• Presentation	12	9.68	Mostly, a presentation was in the form of pair work and groupwork. However, for some participants, each student had one individual presentation at the end of the course.

According to Table 3, fill-in-the-gap tasks and short-answer tasks were the most prevalent, accounting for 37.90% and 31.45% of all items, respectively. Many lecturers said that they mostly used fill-in-the-gap tasks in their listening and vocabulary tasks. For example, Suda said, "*Fill-in-the-gap tasks may assess students' listening ability whether or not they can understand the listening texts. Additionally, such tasks might assess students' vocabulary skills. If they are able to finish the task, they are proficient in vocabulary.*" (Suda). Moreover, some lecturers reported using short-answer tasks to stimulate students' thinking. For instance, Petch stated, "*Once students have mastered the language content, I also encourage them to think by assigning short-answer tasks. This may also provide students with feedback and evaluate their understanding.*" (Petch). Additionally, oral discussions (20.97%) were used mostly in group work to enable students to collaborate and gain knowledge from others. As Chain said, "*I adopted oral discussions among student groups frequently. This is to diagnose and evaluate learner understanding. They may not be confident enough to share with the whole class at first; thus, enabling them to work in groups would increase their confidence.*" (Chain). Finally, presentations (9.68%) were also adopted mostly at the end of the course. In support of this, Mali stated, "*Students have final projects, so they have the opportunity to give presentations. I may assess their comprehension and language proficiency, mostly in the areas of listening and speaking, evaluate their prior learning, and provide them with feedback.*" (Mali). Based on classroom observations, the lecturers' primary responsibilities were that of facilitators. To elaborate, lecturers mostly provided students with cues for their assessment tasks rather than directly helping them. Significantly, most of them adopted assessment practices on digital platforms in which every student could get access to the tasks easily.

According to these findings, the concept of practicality (Díaz et al., 2021) may explain the predominance of the adoption of fill-in-the-gap and short-answer tasks, since they are simple to develop and to grade. In addition, these assessment methods were easy to adopt when evaluating and diagnosing students' learning and understanding. This finding appears to

corroborate Frodden et al.'s (2009) assertion that this practice may be related to teachers' lack of time. Thus, instructors must seek for objective items that are simple to correct and develop rather than subjective tasks. Significantly, less adopted methods are oral discussion and presentation respectively. This appears to contradict what Dunbar et al. (2006) demonstrate, i.e., that one of the most important soft skills that higher education students should acquire prior to graduation is the ability to share information publicly, clearly, and eloquently in accordance with various academic and professional contexts. With this disparity, consequently, Chan (2011) asserts that many undergraduates enter the labor market with severe deficits in oral communication skills in English. To provide some solutions to this issue, it is necessary that students have more opportunities to observe and participate in relatively frequent oral discussion and presentation practices and receive explicit, immediate, and constructive feedback. This might include videotaping the oral presentations (e.g., Dupagne et al., 2007) and using peer assessment strategies (e.g., Murillo-Zamorano & Montanero, 2018) in order to overcome time restrictions.

Student involvement in the assessment task

Since Black and Wiliam's (1998) seminal publication, the effectiveness of formative assessment learning has been well established. Current ideas of formative assessment emphasise student involvement in assessment (Brookhart, 2011; Wiliam, 2011). This research also investigated this aspect of formative assessment by examining the utilisation of self- and peer-assessment in EFL classrooms.

Table 4
Student-involvement in the assessment task

Student-involvement tasks	No. of Tasks	Percentage
• Self-assessment	2	1.61
• Peer-assessment	3	2.42
• Class reflection	9	7.26
• None	110	88.71

According to Table 4, the analysis that categorised each observed task according to the kind of student involvement (self-assessment, peer-assessment, and class reflection) revealed that class reflections accounted for 7.26 percent of all teacher assessments. During 88.71%, most lecturers did not pay attention to any predetermined types of student-involvement, although some of them used self-assessment (1.61%) and peer-assessment (2.42%) in some tasks. These findings on student-involved language assessment mirrored those of prior research (e.g., Alkharusi et al., 2012), indicating that EFL instructors reported developing student-involved assessment less often than other assessment practice. Moreover, digital contexts may pose several challenges; for example, there is less interaction among learners and lecturers in the digital contexts (Madhubhashini, 2022). Contributory explanations that may have led to these findings include: 1) The teaching context promotes teacher-centred instruction and assessment where the teacher is the sole assessor as they might believe that students could not share effective feedback to their peers due to trust and expectation issues (e.g., Carlsson et al., 2022),

2) Teachers may have insufficient training on how to conduct self- and peer-assessment appropriately in the classroom (e.g., Bennet, 2011; Arrafii & Suhaili, 2018), and 3) Teachers may believe that allowing students to engage in self- and peer-assessment could negatively impact classroom management, particularly with regard to time constraints (e.g., Havard et al., 2023). The application of explicit measures, i.e., scoring written reports on a given scale and not merely offering comments and facilitating interpretation of student assessments, is one way to combat certain unfavorable perceptions of self- and peer-assessments. Some instructors and students might also benefit from introductory tutoring sessions that elaborate on the shift from instructor to student assessment responsibility.

Types of teacher feedback given on the task

In this study, feedback is categorised into 1) immediate/non-immediate, 2) written/verbal, and 3) digital/other. According to the observation, most lecturers adopted digital gadgets/applications (e.g., Powerpoint, and Google Docs) to give feedback. Mostly, concerning in-class assessment tasks, they provided immediate feedback right after the end of the task. Suda mentioned that, for instance, "*I like to give them immediate feedback because my students still focus on the task; thus, they could gain more understanding about the assessment task.*" (Suda). This is supported by Brookhart (2011) indicating that instructors should be able to give feedback to their students. In contrast, Mali stated that "*Sometimes, I give students feedback in the next class due to the class time, but I assign them to figure out about the task for their homework. Then in the next class, we could discuss feedback and the task together.*" (Mali). This finding of Mali mirrored the finding of Al Bahlani (2019) emphasising class-time management. Also, lecturers provided students with both written and verbal feedback, which is consistent with Hosseini and Azarnoosh (2014).

All in all, based on digitally-mediated language assessment practice (D-LAP), most lecturers were at the intermediate level in digital literacy (Eyal, 2012) since they mostly involved the administration of examinations, tasks, and projects in a digital environment, with performance indicators created in collaboration with students.

CONCLUSION

As a case study, this study examined the digitally-mediated language assessment practices of four Thai EFL university lecturers. The majority of lecturers demonstrated an intermediate level of digital literacy, as they delivered their assessment tasks in a digital environment, developed performance indicators in collaboration with students, and provided students with feedback. In addition, their assessment tasks were designed based on those in the coursebook in order to maintain the instructors' reliability throughout all course sections. The majority of their assessment duties consisted of evaluating students' learning and diagnosing their knowledge using a range of assessment methods, such as fill-in-the-gap tasks and short-answer tasks.

IMPLICATIONS AND RECOMMENDATIONS

Since the purpose of our research was to explore the digitally-mediated language assessment practices of Thai EFL university lecturers, we obtained data and insights that would enable us to follow this issue in more depth in future large and longitudinal studies. However, there are a few limitations that must be considered. First, cautions should be taken when interpreting the results, since we relied only on case studies from four lecturers. Due to the qualitative nature of our research, we were unable to locate a large sample that is representative of the population of Thai EFL university lecturers. This limits the generalizability of our results; nevertheless, this study enabled readers to zoom in on the real-world context of language classrooms in order to transfer how Thai EFL university lecturers test and assess their students to their own contexts through the concept of 'transferability', allowing readers to make connections between the findings of this study and their own experiences and contexts. Case study research with sufficient information to obtain context-specific findings and discern linkages between individual factors, personal experiences, and social and institutional contexts may be more appropriate for addressing the issue. In addition, more incorporation from other stakeholders such as university policy makers, deans, or course developers is needed as they might be able to provide more insightful information in relation to language assessment literacy of lecturers. Furthermore, additional multiple assessment measures, such as an evaluation of instructor-created assessment task documents, could provide a more comprehensive view of instructors' language assessment, thereby fostering more targeted language assessment professional development programs.

Declaration of conflicting interest

The authors declare that there is no conflict of interest in this work.

THE AUTHORS

Pariwat Imsa-ard is a PhD candidate in English Language Teaching at Language Institute, Thammasat University, Thailand. He has an M.A. in TESOL from the University of York. His research interests include language assessment, English language teaching, reflective practice, and interculturality.
pimsaard@gmail.com

Supong Tangkiengsirisin, PhD, is an associate professor at Language Institute, Thammasat University, Thailand. He has a PhD from the University of Nottingham, UK. His research interests include second language writing, written discourse analysis, and English for specific purposes (ESP).
supong.t@litu.tu.ac.th

REFERENCES

Al-Bahlani, S. M. (2019). *Assessment literacy: A study of EFL teachers' assessment knowledge, perspectives, and classroom behaviors* [Doctoral dissertation, The University of Arizona]. UA Campus Repository. https://repository.arizona.edu/handle/10150/633240

Alhat, S. (2020). Virtual classroom: A future of education Post-COVID-19. *Shanlax International Journal of Education, 8*(4), 101-104. https://doi.org/10.34293/education.v8i4.3238

Alkharusi, H. (2021). English language teachers' uses of classroom assessment. *Anatolian Journal of Education, 6*(1), 103-108. https://doi.org/10.29333/aje.2021.618a

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning, 1*(2), 217-232. http://doi.org/10.5539/jel.v1n2p217

Arksey, H., & Knight, P. T. (1999). *Interviewing for social scientists: An introductory resource with examples*. Sage.

Arrafii, M. A., & Sumarni, B. (2018). Teachers' understanding of formative assessment. *Lingua Cultura, 12*(1), 45-52. https://doi.org/10.21512/lc.v12i1.2113

Astiandani, F. R., & Anam, S. (2021). EFL teachers' perceptions towards the implementation of online formative assessment amidst the COVID-19 pandemic. *ELT Worldwide: Journal of English Language Teaching, 8*(2), 269-277. https://doi.org/10.26858/eltww.v8i2.21326

Babaii, E., & Asadnia, F. (2019). A long walk to language assessment literacy: EFL teachers' reflection on language assessment research and practice. *Reflective Practice, 20*(6), 745-760. https://doi.org/10.1080/14623943.2019.1688779

Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal, 73*(2), 113-123. https://doi.org/10.1093/elt/ccy055

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25. https://doi.org/10.1080/0969594X.2010.513678

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3-12. https://doi.org/10.1111/j.1745-3992.2010.00195.x

Brown, A. (2013). Multicompetence and second language assessment. *Language Assessment Quarterly, 10*(2), 219-235. https://doi.org/10.1080/15434303.2013.769551

Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson.

Bryman, A. (2008). *Social research methods* (3rd ed.). Oxford University Press.

Carlsson Hauff, J., & Nilsson, J. (2022). Students' experience of making and receiving peer assessment: The effect of self-assessed knowledge and trust. *Assessment & Evaluation in Higher Education, 47*(6), 959-971. https://doi.org/10.1080/02602938.2021.1970713

Chan, V. (2011). Teaching oral communication in undergraduate science: Are we doing Enough and doing it right? *Journal of Learning Design, 4*(3), 71-79.

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing, 21*(3), 360-389. https://doi.org/10.1191/0265532204lt288oa

Chung, S. J., & Choi, L. J. (2021). The development of sustainable assessment during the COVID-19 pandemic: The case of the English language program in South Korea. *Sustainability, 13*(8), 4499. https://doi.org/10.3390/su13084499

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed-methods approaches*. Sage.

DeLuca, C., & Johnson, S. (2017). Developing assessment capable teachers in this age of accountability. *Assessment in Education: Principles, Policy & Practice, 24*(2), 121-126. https://doi.org/10.1080/0969594X.2017.1297010

Díaz, C., Jara, A., Rosales, Y., & Sanhueza, M. (2021). Characterizing english assessment instruments: An overview of their design. *Revista Chakiñan de Ciencias Sociales y Humanidades*, (15), 80-96. https://doi.org/10.37135/chk.002.15.05

Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education, 31*, 115-128. https://doi.org/10.1007/s10755-006-9012-x

Dupagne, M., Stacks, D. W., & Giroux, V. M. (2007). Effects of video streaming technology on public speaking students' communication apprehension and competence. *Journal of Educational Technology Systems, 35*(4), 479-490. https://doi.org/10.2190/5947-4W72-303L-Q578

Duraku, Z. H., & Hoxha, L. (2020). The impact of COVID-19 on education and on the well-being of teachers, parents, and students: Challenges related to remote (online) learning and opportunities for advancing the quality of education. In Z. H. Duraka (Ed.), *Impact of the COVID-19 pandemic on education and wellbeing: Implications for practice and lessons for the future* (pp. 17-45). University of Prishtina Hasan Prishtina.

Eyal, L. (2012). Digital assessment literacy—The core role of the teacher in a digital environment. *Journal of Educational Technology & Society, 15*(2), 37-49.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113-132. https://doi.org/10.1080/15434303.2011.642041

Frodden, M. C., Restrepo, M. I., & Maturana, L. (2009). Analysis of assessment instruments used in foreign language teaching. *Íkala, Revista de Lenguaje y Cultura, 9*(1), 171-201. https://revistas.udea.edu.co/index.php/ikala/article/view/3146

Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. Routledge.

Havard, B., Podsiad, M., & Valaitis, K. (2023). Peer assessment collaboration evaluation: An innovative assessment tool for online learning environments. *TechTrends*, 1-11. https://doi.org/10.1007/s11528-022-00832-8

Hosseini, A. S., & Azarnoosh, M. (2014). Iranian EFL instructor's oral assessment practices: Purposes, methods, procedures. *Procedia-Social and Behavioral Sciences, 98*, 653-658. https://doi.org/10.1016/j.sbspro.2014.03.464

Imsa-ard, P. (2020). Voices from Thai EFL teachers: Perceptions and beliefs towards the English test in the national examination in Thailand. *LEARN Journal: Language Education and Acquisition Research Network, 13*(2), 269-287.

Imsa-ard, P. (2023). *Language assessment literacy: Exploring Thai EFL teachers' assessment knowledge, beliefs, and digitally-mediated assessment practices in classrooms* [Doctoral dissertation]. Thammasat University.

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/10.1177/0265532208090158

Jan-nesar, M. Q., Khodabakhshzadeh, H., & Motallebzadeh, K. (2020). Assessment literacy of Iranian EFL teachers: A review of recent studies. *Journal of Asia TEFL, 17*(2), 689-698. https://doi.org/10.18823/asiatefl.2020.17.2.27.689

Kantabutra, S., & Tang, J. (2010). Efficiency analysis of public universities in Thailand. *Tertiary Education and Management, 16*(1), 15-33. https://doi.org/10.1080/13583881003629798

Kaur, A., Young, D., & Kirkpatrick, R. (2016). *English language education policy in Asia*. Springer. https://doi.org/10.1007/978-3-319-22464-0_16

Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication Methods and Measures, 5*(2), 93-112. https://doi.org/10.1080/19312458.2011.568376

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169-197. https://doi.org/10.1177/0265532214554321

Liamputtong, P. (2009). Qualitative data analysis: conceptual and practical considerations. *Health Promotion Journal of Australia, 20*(2), 133-139. https://doi.org/10.1071/HE09133

Madhubhashini, G. T. (2022). Digital divide: Challenges faced by the university students in the context of Open and Distance Learning (ODL) during the covid-19 pandemic. *The Journal of Development Communication, 33*(1), 16-23. http://jdc.journals.unisel.edu.my/index.php/jdc/article/view/222

Murillo-Zamorano, L. R., & Montanero, M. (2018). Oral presentations in higher education: A comparison of the impact of peer and teacher feedback. *Assessment & Evaluation in Higher Education, 43*(1), 138-150. https://doi.org/10.1080/02602938.2017.1303032

Nam, Y. (2021). *Exploring language teachers' assessment practices and language learners' perceptions towards assessment in the online environment* [Doctoral dissertation, Georgia State University] ScholarWorks @Georgia State University. https://doi.org/10.57709/26839572

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*, 533-544. https://doi.org/10.1007/s10488-013-0528-y

Patton, M. Q. (2002). Two decades of developments in qualitative inquiry: A personal, experiential perspective. *Qualitative Social Work, 1*(3), 261-283. https://doi.org/10.1177/1473325002001003636

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice, 48*(1), 4-11. https://doi.org/10.1080/00405840802577536

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator, 46*(4), 265-273. https://doi.org/10.1080/08878730.2011.605048

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers* (2nd ed.). Blackwell Publishing.

Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education, 17*(1), 1-24. https://doi.org/10.1207/s15324818ame1701_1

Smith, L. F., Hill, M. F., Cowie, B., & Gilmore, A. (2014). *Designing assessment for quality learning*. Springer.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*(3), 238-246.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2006). *Classroom assessment for student learning: Doing it right, using it well*. Assessment Training Institute.

Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia, 9*(1), 1-14. https://doi.org/10.1186/s40468-019-0077-8

Taylor, L., & Harding, L. (n.d.). *A testing time for testing: Assessment literacy as a force for social good in the time of coronavirus*. https://acss.org.uk/a-testing-time-for-testing-assessment-literacy-as-a-force-for- social-good-in-the-time-of-coronavirus/

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics, 29*, 21-36. https://doi.org/10.1017/S0267190509090035

van der Lans, R. M., van de Grift, W. J., van Veen, K., & Fokkens-Bruinsma, M. (2016). Once is not enough: Establishing reliability criteria for feedback and evaluation decisions based on classroom observations. *Studies in Educational Evaluation, 50*, 88-95. https://doi.org/10.1016/j.stueduc.2016.08.001

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 11*(4), 374-402. https://doi.org/10.1080/15434303.2014.960046

Watmani, R., Asadollahfam, H., & Behin, B. (2020). Demystifying language assessment literacy among high school teachers of English as a foreign language in Iran: Implications for teacher education reforms. *International Journal of Language Testing, 10*(2), 129-144.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14. https://doi.org/10.1016/j.stueduc.2011.03.001

Xie, Q., & Tan, S. (2019). Preparing primary English teachers in Hong Kong: Focusing on language assessment literacy. *Journal of Asia TEFL, 16*(2), 653-673. https://doi.org/10.18823/asiatefl.2019.16.2.14.653

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.

Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia-Social and Behavioral Sciences, 116*, 2998-3004. https://doi.org/10.1016/j.sbspro.2014.01.696

Yin, R. K. (1994). Discovering the future of the case study. Method in evaluation research. *Evaluation Practice, 15*(3), 283-290. https://doi.org/10.1177/109821409401500309

Appendix A

An observation classroom schedule

Observation Checklists	Describing Tasks	Notes
Source of the assessment task • Teacher developed • Coursebook • Other		
Purposes of the Assessment Tasks • Diagnosing • Grading • Evaluating • Reporting feedback • Improving instruction • Other		
Assessment Methods • MCQs • True-False • Fill-in-the-gap • Short answers • Oral discussion • Presentation • Other		
Student-Involvement in the Assessment Tasks • Self-assessment • Peer-assessment • Class reflection • None • Other		
Teachers' Feedback • Immediate / Non-Immediate • Written / Verbal • Digital • Other		
Role of the Teacher • Directly helping students • Giving cues • Other		

Appendix B

A stimulated recall protocol

Potential questions for stimulated recalls:

1. What is the purpose of this particular assessment task? (design)
2. Did you develop this assessment task yourself? Whys so? (design)
3. What is the purpose behind making this assessment task? (delivery)
4. Did the assessment task succeed in achieving the purpose? How do you know? (decision-making)
5. If you have the opportunity to redesign this assessment task, what would you do? Why? (evaluating the task).