PASAA Journal
Volume 66, October 2023, 37-80

Language Assessment Literacy: A Preliminary Survey of Thai EFL University Lecturers' Language Assessment Knowledge, Beliefs, and Practices

Pariwat Imsa-ard[a]* and Supong Tangkiengsirisin[a]

[a] Language Institute Thammasat University, Bangkok, Thailand

**Corresponding author: pimsaard@gmail.com*

Article information	
Abstract	Language assessment literacy (LAL) remains an uncharted area for many English language lecturers, hindering their ability to effectively bridge the gap between theoretical knowledge and practical application in language assessment within the classroom setting. This research aimed to investigate the language assessment literacy of Thai EFL university lecturers, with a focus on the reciprocal relationship between their knowledge and beliefs and their assessment practices. Employing a multistage sampling method, the study involved 152 Thai EFL university lecturers from eight public universities in Thailand. Participants completed a language assessment literacy questionnaire and assessment practice inventory. The findings indicated that the majority of Thai EFL university lecturers demonstrated a moderate level of language assessment knowledge. Notably, they exhibited higher proficiency in articulating clear learning objectives but lower proficiency in providing effective feedback. Moreover, the lecturers expressed that performance assessment held the greatest significance in classroom assessment, while traditional assessment methods and

E-ISSN: 2287-0024

	non-achievement-based grading were deemed less important. Additionally, it was revealed that lecturers often communicated grading criteria in advance, but they paid little attention to item analysis and test revision based on its analysis. These findings are expected to contribute to the understanding of English language lecturers' language assessment literacy and provide a basis for further development in this area.
Keywords	language assessment literacy, assessment knowledge, beliefs, practices, Thai EFL university lecturers
APA citation:	Imsa-ard, P. & Tangkiengsirisin, S. (2023). Language assessment literacy: A preliminary survey of Thai EFL university lecturers' language assessment knowledge, beliefs, and practices [Special issue]. *PASAA, 66*, 33–76.

1. Introduction

English language Instructors are key players in language classroom. Among their many responsibilities, instructors hold a crucial role in assessing their students' learning performance within the classroom. Notably, the work of Black and Wiliam (1998) has highlighted the importance of assessment in promoting student learning, leading to increased attention to language assessment within the classroom (Fulcher, 2012). Consequently, instructors are expected to possess comprehensive understanding and experience in various aspects of assessment, including assessment literacy, to effectively accomplish their assessment objectives. Recognizing the significance of this matter, the language assessment literacy of instructors is paramount in enhancing student learning outcomes (Fulcher, 2012). Furthermore, there exists a consensus that instructors' knowledge and beliefs regarding assessment can significantly influence their practices in classroom assessment (Popham, 2009). Empirical research (e.g., Lam, 2015; Watmani et al., 2020) has consistently demonstrated inadequate testing and assessment knowledge and skills among language instructors. Moreover, previous studies (e.g., DeLuca & Johnson, 2017; Yamtim & Wongwanich, 2014) have highlighted the lack of assessment knowledge among instructors, particularly in terms of its practical application, confirming Vogt and Tsagari's (2014) assertion that some instructors possess insufficient knowledge of language assessment terminology for effective implementation. Furthermore, the pervasive influence of an exam-oriented culture in Thailand (Kaur et al., 2016) reinforces the focus on testing and assessment practices that prioritize test-taking strategies at the expense of broader language skills. In the Thai context, Imsa-ard (2020, 2021) has noted that Thai EFL instructors often resort to teaching to the test in order to help students achieve higher test scores, thereby limiting the curriculum's scope and resulting in negative washback effects. Since the practices have been repeatedly adopted and shaped by their teaching experience, this possibly leads to the fact that instructors might not seek and adopt innovative assessment practices that might enhance the quality of instruction. These circumstances underscore the

evident gap between assessment principles and their practical application, with many instructors lacking formal training in assessment (Berry et al., 2019). Consequently, their assessment practices may be shaped by their own past testing and assessment experiences, as observed by Smith et al. (2014) and Vogt and Tsagari (2014). To support this, several scholars (e.g., Csépes, 2021; Tsagari & Vogt, 2017; Xu & Liu, 2009) also argue that past experiences can play a decisive role in the development of their assessment literacy as it can account for the fact that they favor well-known, mostly traditional assessment methods without accepting effective and innovative methods. Remarkably, research studies by Barnes et al. (2014) and Brown and Harris (2009) have revealed that instructors' assessment beliefs and practices are also influenced by school regulations as well as societal, cultural, and social values. Furthermore, Thong-Iam and Subphadoongchone (2019) suggest that university lecturers require training to effectively implement assessment practices in the classroom, as successful assessment practices demand the development of assessment-related knowledge and skills alongside the application of appropriate principles and conceptions.

The assessment literacy of instructors has been identified as an underdeveloped area in various research studies (e.g., Jan-nesar et al., 2020; Sultana, 2019; Xu & Brown, 2017), highlighting the need for further investigation into the assessment literacy of instructors across different contexts. Additionally, the emergence of the COVID-19 pandemic has raised significant assessment-related concerns among students, parents, and educators (Duraku & Hoxa, 2020). To elaborate, these stakeholders were concerned with an inadequate level of readiness and support to implement effective and successful instruction and assessment. In response to these concerns, Astiandani and Anam (2021) have discovered that instructors require support in dealing with technological challenges and adapting to the use of digital devices for assessment purposes. These concerns, coupled with existing disparities in assessment practices, pose obstacles to effective assessment practices aimed at enhancing student learning

outcomes. Therefore, it is essential to examine the language assessment literacy and assessment practices of instructors to gain a better understanding by addressing the following research questions:

1. To what extent do Thai EFL lecturers possess language assessment knowledge?
2. What are Thai EFL lecturers' beliefs towards their assessment literacy?
3. What are classroom assessment practices of Thai EFL lecturers?

2. Literature Review

2.1 Language Assessment Literacy

Without a doubt, testing and assessment continue to increase in significance and impact across the globe today. Assessment takes up a significant portion of teaching and learning time. In recent times, there has been an increasing focus on enhancing instructors' language assessment literacy (Coombe et al., 2020) which is critical to increasing students' learning and to the success of teaching (Fulcher, 2012; Wang et al., 2008). Assessment literacy, according to Stiggins (1991) who firstly coined this term based on the assessment competence standards, refers to the knowledge, skills, and concepts of sound assessment that stakeholders involved in assessment practices need to acquire. When it comes to being deemed assessment literate, as Stiggins (1999) has demonstrated, instructors must be able to link assessments to specific purposes, express expectations for students' achievement, utilize appropriate assessment techniques, and set quality scoring standards in order to minimize bias in assessment. Popham (2011) has defined assessment literacy as "an individual understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions" (p. 267), which implies that instructors should be able to make inferences and draw conclusions about students based on their behavioral repertoire as well as their covert knowledge and skills. In addition, similar to assessment literacy in general, Brookhart (2011) provides an updated list of knowledge and skills instructors should possess in order to conduct the

assessment-related tasks competently and professionally. Such lists are specifically updated to include knowledge and skills for formative assessment. More precisely, this work also offers educational assessment knowledge and skills for instructors that align with current instructor assessment standards.

When drawing from general assessment literacy, language assessment literacy refers to stakeholders' knowledge of assessment practices, as well as their application of this knowledge to classroom activities in general, and especially to concerns of language assessment (Taylor, 2009). Language assessment literacy has leaned extensively on general assessment literacy literature and research, while also attempting to distinguish itself as a knowledge base that incorporates unique elements inherent in conceptualizing and measuring language-related performance. Furthermore, several definitions of language assessment literacy have been offered in light of expanding data on assessment needs and practices. O'Loughlin (2013), for instance, defines language assessment literacy as something that:

> *"potentially includes the acquisition of a range of skills related to test production, test score interpretation and use, and test evaluation in conjunction with the development of a critical understanding about the roles and functions of assessment within education and society"* (p. 363).

Due to the COVID-19 pandemic, face-to-face classrooms have been shifted to virtual classrooms, which forced instructors to adopt digital applications and gadgets to facilitate their teaching and assessment. Instructors who are digitally literate should be able to utilize a number of applications and technological systems in order to develop skills of students, adopting a variety of assessment methods. Eyal (2012) focuses on a distinct facet of assessment literacy—digital assessment literacy. The term "digital assessment literacy" was not used in academic literature prior to the year 2021; however, the term alludes to the

instructors' role as assessors in a technology-rich context. Eyal (2012) has then proposed the following abilities and skills that instructors must possess in order to achieve digital assessment literacy on a three-tiered continuum: basic, intermediate, and advanced digital assessment literacy. As with other types of literacy, digital assessment literacy develops with interaction with the environment, which includes not just the human environment but also the digital world and its potential. Instructors working in a digital environment must acquire information and skills that will aid them in choosing and using digital assessment tools.

In this study, the focus of the investigation was on university lecturers. In light of this, 'language assessment literacy' in this study has been operationally defined as language instructors' knowledge, beliefs, and digitally-mediated assessment practices. Over the last few decades, researchers (e.g., Coombe et al., 2020; Fulcher, 2012; Lam, 2015; Popham, 2006; Stiggins, 1995) have critiqued weak instructor competency in assessment literacy. To date, a slew of studies on instructor assessment literacy have concentrated on instructors' assessment knowledge and perspectives on language assessment literacy (Coombe et al., 2020). This is hardly unexpected given the widespread belief that instructors' assessment knowledge and beliefs influence their classroom practices (Alkharusi et al., 2012; Popham, 2009; Sultana, 2019).

2.2 Assessment Beliefs and Practices

Despite the fact that the term "belief" is contested, a number of scholars have attempted to define it to date. Some research on instructors' beliefs has a longer history of study and has more clearly defined the constructs of knowledge and belief. However, researchers studying instructors' assessment beliefs use a variety of subsuming terminology to characterize the variables of interest, including "conceptions" (as defined by Thompson, 1992). Conceptions, according to Thompson (1992), are part of "a broader mental framework that comprises beliefs,

meanings, ideas, propositions, rules, mental representations, and preferences" (p. 130). In essence, a conception is a single entity that condenses knowledge and belief and serves as a framework for describing instructors' overall perception and awareness of assessment. Additionally, concerning beliefs about the different forms of assessment, instructors have varied beliefs about the effectiveness of various assessment methods, such as formative assessment and standardized tests. Furthermore, Brown and Harris (2009) suggest that instructors' assessment beliefs vary by context, reflecting their absorption of society's cultural purposes and practices. Regarding language assessment, instructors' belief in their test designs and test administration can be in direct conflict with the assessment purposes of the curriculum, which focuses more on building language skills in communication. Additionally, instructors' beliefs that assessing students' understanding of linguistic items may help them prepare adequately for final exams imply a considerable influence of high-stakes tests on their classroom assessment practices (Ha et al., 2021; Imsa-ard, 2020). These ideas run counter to the ideals of innovative language teaching principles (Bachman, 2002; Norris, 2016) and the teaching curriculum. These beliefs may have been shaped by instructors' language learning experiences in which they were taught via the traditional grammar-translation method, as well as by their own prior teaching experiences (Ha & Murray, 2021).

Concerning assessment practices, although Black and Wiliam (1998) have demonstrated that classroom assessment is commonly understood to include a wide range of activities that instructors and learners engage in to gather data that may be utilized diagnostically to enhance teaching and learning, classroom assessment practices are widely assumed to be multifaceted and multidimensional (Rodriguez, 2004). Stiggins et al. (2006) have proposed five indicators of sound classroom assessment practices as follows: "1) clear purposes, 2) clear targets, 3) sound design, 4) effective communication, and 5) student involvement" (p. 27). To support Stiggins et al. (2006) about student involvement,

Black and William (2009) have demonstrated that innovative language learning theories emphasize how students learn depending on formative assessment information they receive and its interpretation, allowing them to make successful learning choices. With this in mind, student involvement in the use of assessment information necessitates an expansion of instructors' assessment knowledge and skills, on account of the fact that this will require instructors to be adept at articulating and sharing learning intentions and criteria for success with their students, as well as offering meaningful opportunities for students to act on assessment information they receive.

When it comes to assessment practices, in addition, the COVID-19 pandemic could mark a watershed moment in which instructors can implement full online teaching and assessment at all times, allowing them to accommodate large classes while also making teaching and assessment more appropriate for the young generation, known as digital natives (Alghammas, 2020). However, there is an evident lack of research (e.g., Jeong, 2013; Noijons, 2013; Wang, 2014) that investigates knowledge, beliefs, and skills of language instructors when it comes to using technology for language testing in their virtual classrooms, which refers to the use of computers in testing as part of digitally-mediated assessment literacy. Significantly, more investigations into beliefs and practices should be done to bridge the gap between them.

2.3 Related Studies

It is widely accepted that language instructors with a good assessment background may integrate assessment into lessons to apply the most successful teaching methods. Despite the high regard of assessment literacy, progress has been slow. While there is a wealth of literature and research on language instructors' assessment practices in an EFL context (e.g., Cheng et al., 2004; Cheng et al., 2008; Inbar-Lourie & Donits-Schmidt, 2009; Narathakoon et al., 2020; Qian, 2008), assessment literacy and instructor knowledge are lacking, and some

Thai secondary school teachers claimed little or no assessment training (Imsa-ard, 2020, 2021).

To begin with, Vogt and Tsagari (2014) conducted a study to determine the training needs of "regular" foreign language instructors, their knowledge of language testing and assessment (LTA), and the extent to which they received training in these areas during their pre-service and in-service education. Using a questionnaire (N = 853) and interviews (N = 63), they surveyed 853 language instructors from seven European countries (Cyprus, Germany, Greece, Italy, Poland, and Turkey) regarding their language testing and assessment skills, as well as their individual opinions regarding their training in Europe. According to the study findings, the majority of language instructors had "insufficient" or "no" training in language testing and assessment. More recently, Watmani et al. (2020) examined the assessment literacy of 200 Iranian EFL high school teachers in order to enhance teacher education. They used a quantitative method to investigate teachers' assessment literacy in connection with the seven standards for language teacher competence while assessing students' learning success. According to the findings, participants lacked comprehension of assessment literacy principles and procedures. They found that EFL teachers with TEFL experience performed better on literacy tests than those without. However, in their study, there was a need to explore how EFL teachers used student-involved assessment such as peer-assessment in their language assessment. In Thailand, there were some studies related to language assessment literacy that should be noted. For example, Viengsang (2016) attempted to explore the assessment literacy of pre-service teachers doing their practicum at public and private secondary schools in Bangkok, Thailand, using an assessment literacy survey questionnaire and a semi-structured interview protocol. Her findings suggested that although participants had undergone training that enabled them to acquire some language assessment knowledge, they seemed unable to apply it in their classroom due to some

misconceptions. However, it should be noted that her study contributed to the literature yet highlighted some participant gaps that this study attempted to bridge.

Concerning language assessment beliefs and practices, Narathakoon et al. (2020) investigated beliefs about classroom assessment in English, their actual classroom practices, and the extent to which their beliefs aligned with their actual assessment practices of sixth-grade English teachers in a school district in Northeastern Thailand. They employed a questionnaire, observation, stimulated recall, and interviews to explain and support teachers' beliefs and assessment practices in the classroom. A lack of assessment training may also be detrimental to assessment practices, suggesting that professional development was vital for teachers' beliefs and classroom assessment methods. This conclusion was supported by some research (e.g., Acar-Erdol & Yildizli, 2018; Hussain et al., 2019; Imsa-ard, 2020), indicating that teachers had trouble transferring their beliefs into the real classroom, since many acknowledged a lack of assessment knowledge. To address this issue, researchers suggested that language instructors should have both theoretical and practical knowledge of language assessment to implement in their classrooms; hence, it is necessary to assess their language assessment knowledge. Moreover, Bashitialshaaer et al. (2021), in light of the COVID-19 pandemic, sought to identify and understand the obstacles and barriers to successfully implementing electronic exams in distance education by conducting an exploratory descriptive study with a sample of 152 university instructors and students from four of Gaza's largest universities. The findings showed that university professors and students encountered many challenges, which were consistent with the literature and may be addressed by improved test design, training, and preparation, or the use of appropriate software.

According to the literature review, there have been a number of studies on language assessment literacy from the perspective of instructors in various teaching levels and contexts such as the study focusing on primary school teachers

(Xie & Tan, 2019) and studies focusing on Israel or European contexts (Levi & Inbar-Lourie, 2019; Watmani et al., 2020). Previous research has conducted limited investigations into the language assessment literacy, encompassing knowledge, beliefs, and assessment practices of instructors, with no specific focus on Thai EFL university lecturers. This creates an obvious research gap that was addressed in this study. The purpose of the present research was to explore the current levels of language assessment literacy among Thai EFL university lecturers, including their knowledge, beliefs, and assessment practices in the classroom.

3. Research Methodology

3.1 Research Design

This quantitative survey research aimed to elicit Thai EFL university lecturers' assessment literacy, encompassing knowledge, beliefs, and practices. The survey consisted of three parts: 1) assessment knowledge test focusing on knowledge on language assessment, 2) belief questionnaire focusing on their beliefs about language assessment, and 3) assessment practice inventory focusing on their common assessment practices.

3.2 Participants

This research focused on Thai EFL university lecturers teaching English foundation courses to undergraduate students. The study employed a multistage sampling technique, which involved systematically reducing a large sample size to effectively manage dispersed populations. The sample size was determined using *G*Power software* (Kang, 2021), considering an effect size of .32, an alpha (α) of .05, and a power of .95, based on relevant studies (e.g., Alkharusi, 2011; Cohen, 1988; Wiratchai, 2012). At least 111 participants were required, selected from eight universities ranked in both QS World University Rankings 2022 and SCImago Institutions Rankings to ensure high-quality instruction in higher education. A total of 152 university lecturers volunteered to participate in this study, surpassing the expected number.

As seen in Table 1, demographic information of the participants is demonstrated. The bulk of participants were female (52.63%), while 47.37% were male. The majority of them were between the ages of 31 and 35 (36.84%) and 36 and 40 (26.32%). They had been teaching English for 11 to 15 years and six to ten years (36.85% and 31.58%, respectively). Almost two-thirds of them (63.16%) held a doctoral degree, while 36.84% graduated with a Master's degree. Lastly, fewer than one-third (31.58%) taught three courses per semester.

Table 1

Demographic Information of Participants in this Study

Criteria		Percentage
Gender	Male	47.37
	Female	52.63
Age range	20-30	15.79
	31-35	36.84
	36-40	26.32
	41-45	10.53
	46-50	5.26
	51 or more	5.26
Years of teaching	1-5	21.05
	6-10	31.58
	11-15	36.84
	16-20	10.53
	21 or more	-
Highest educational qualification	Doctorate	63.16
	Master's	36.84
Number of courses taught	1	-
	2	26.32%
	3	31.58%
	4	26.32%
	5 or more	15.79%

3.3 Research Instruments

3.3.1 Assessment Knowledge Test

This test consisted of 20 multiple-choice questions designed based on Brookhart's (2011) list of Educational Assessment Knowledge and Skills for Teachers and Eyal's (2012) work on digitally-mediated assessment literacy. The themes were as follows: 1) understanding language learning and teaching, 2) articulating clear learning intentions, 3) communicating learning outcomes, 4) using test methods, purposes, validity, fairness, and reliability, 5) analyzing test items, 6) communicating effective feedback, 7) scoring, 8) result-informed decision-making, 9) communicating with other stakeholders, 10) helping students plan their learning, 11) preparing tests, and 12) digitally-mediated assessment literacy. In developing the test items, the participants' work contexts (e.g., where, who, and what they taught) were also taken into account. Consideration was given to the amount of control lecturers had over assessment practices and the assessment materials they employed. In addition, Google Forms was used to administer the test online to ensure the participants' convenience. It took each participant approximately 30 minutes to complete the test.

The content validity of the test was measured by the content validity index (CVI) and Kappa statistics. The validation process in this study involved three experts who had a minimum of five years of experience in English language teaching and language assessment. Since CVI does not account for the inflated results that may occur from chance agreement, the Kappa Statistic coefficient was used to ensure a better understanding of content validity. The Kappa statistic is a consensus indicator of interrater agreement that supports the CVI to guarantee that the agreement among experts is not due to chance. The assessment knowledge test had a CVI value of .93, which is considered 'acceptable,' and a Kappa statistic value was .89, which is considered 'excellent.'

3.3.2 A Beliefs Questionnaire

The second data collection instrument was the assessment beliefs questionnaire, which consisted of two parts: background information and assessment beliefs. First, data regarding the participants' gender, age, education, teaching experience, teaching load, and assessment training were collected. The second part of the questionnaire was adapted from the instruments used by Al-Bahlani (2019), Alkharusi (2009), and Alkharusi et al. (2012) to assess instructors' beliefs and perceptions of implementing specific educational assessment tasks in the following categories: 1) developing and administering assessment methods, 2) developing and scoring performance assessment, 3) developing grading procedures, 4) communicating assessment results to various audiences, 5) reviewing assessment results, and 6) using digitally-mediated assessment. In this part, the participants rated the significance of classroom assessment practices on a 5-point Likert scale from 1 (not important) to 5 (very important). Participants' assessment beliefs were determined by averaging all questions in a category.

A predictive validity analysis conducted on the questionnaire's items indicated that the variables were shown to be positively related $r(209) = .41$, $p < .001$, and Cronbach's alpha coefficient for reliability was also found to be .97, which is considered excellent (Bryman, 2008). To establish the validity of the questionnaire, three experts in the field of language assessment were asked to quantitatively and qualitatively evaluate the instruments. The content validity of the beliefs questionnaire was quantified using the index of item objective congruence (IOC). All items had an IOC value of more than .67, indicating acceptable levels of construct and content validity. The experts also suggested rewording several parts of the beliefs questionnaire. In the item statement, for instance, one expert advised

replacing ‘best’ with ‘appropriate’ and ‘develop an answer sheet’ with ‘proposed/potential answers.’

3.3.3 Assessment Practice Rating Scales

First created by Zhang and Burry-Stock (1994) and later revised by Zhang and Burry-Stock (2003), the Assessment Practices Inventory was adopted and adapted in this study. This inventory was developed within the theoretical framework created by the classroom assessment literature (e.g., Airasian, 2000) and the Standards for Teacher Competence in Educational Assessment of Students. The original 67 questions in this inventory cover a broad range of assessment tasks and practices, including creating paper-and-pencil tests and performance measures, interpreting standardized test scores, grading, communicating assessment results, and applying assessment results to decision-making. In this study, university lecturers were requested to rate their responses to the use scale’s items. The use scale is intended to measure language instructors’ assessment practices from the following scale: 1 (never used), 2 (seldom used), 3 (sometimes used), 4 (often used), and 5 (used very often).

This rating scale was chosen because its reliability was confirmed by a Cronbach alpha value of .97, and its construct validity was further assessed using the Rasch model and component analyses. To establish the validity of an assessment practice inventory, three experts in the field of language assessment were asked to quantitatively and qualitatively evaluate the instruments. The content and construct validity of an assessment practice inventory was quantified using the index of item objective congruence (IOC). All items had an IOC value of more than .95, indicating acceptable levels of construct and content validity.

3.3.4 Pilot Study and Its Reliability

After considering the comments and suggestions of the experts, the instruments (i.e., language assessment knowledge test, beliefs questionnaire, and assessment practice rating scales) were piloted with 18 participants whose characteristics were similar to those of the study's target population. According to Connelly (2008), existing research advises that a pilot study sample should consist of 10 percent of the number of participants anticipated for the real study. In addition, pilot research sample size guidelines (e.g., Birkett & Day, 1994; Julious, 2005) recommend at least ten participants; hence, 18 individuals in this study were feasible and sufficient for a pilot study. As for the posteriori validity evidence regarding validity and reliability, an item analysis was conducted. The findings produced strong item discrimination index values, with an average of .41. This indicates that the test may successfully distinguish between degrees of mastery within the language assessment domain. Moreover, Cronbach's alpha coefficient for internal consistency reliability of the beliefs questionnaire was .92, which was considered ideal (Bryman, 2008). In addition, for the assessment practice rating scales, Cronbach's alpha coefficient for internal consistency reliability was .94, which was also considered ideal (Bryman, 2008).

3.4 Data Collection and Analysis

The data collection process began in March 2022. In this quantitative part, data were collected from the assessment knowledge test, teacher assessment beliefs questionnaire, and assessment practice inventory. This took place during March-June 2022 (i.e., the second semester and summer semester of the academic year 2021). As this research involved people, their rights were a crucial issue. Prior to data collection, participants were provided with information sheets in their native language outlining the study's purposes and their ethical rights.

The collected data were analyzed by employing different statistics that could elicit the results and were appropriate to each type of instruments. First, descriptive statistics (mean, *SD*, and percentage) were used to answer questions for the overall participants by grouping them into three domains of interests: knowledge, beliefs, and practices. More specifically, data obtained from Likert scales were interpreted according to Table 2. In order to explore the relationship between language assessment knowledge, belief, and practice, a Pearson correlation was used to determine the correlation between (1) language assessment knowledge and belief, and (2) language assessment knoweldge and practice. In addition, a Multivariate Analysis of Variance (MANOVA) was employed to determine whether teachers significantly differed in their levels of belief and practice across different language assessment knowledge groups. Moreover, pair-wise comparisons from MANOVA were conducted to help work out the importance of years of teaching relative to assessment practices.

Table 2

Interpretations of Likert-scale Surveys (Pallant, 2005)

Ranges	Beliefs	Practices
1.00 – 1.80	unimportant	not at all used
1.81 – 2.60	little important	seldom used
2.61 – 3.40	undecided	used occasionally
3.41 – 4.20	important	used often
4.21 – 5.00	very important	used very often

4. Results

This research study investigated Thai EFL university lecturers' language assessment literacy through surveys of lecturers' language assessment knowledge, beliefs about language assessment, and perceived assessment practices. To contextualize the data, survey items were organized according to these categories. In this section, data collected from a multiple-choice test and Likert scale items for the whole group within each domain are discussed.

4.1 Language Assessment Knowledge

To address the first research question, this study examined Thai EFL university lecturers' performance on a test that measured their knowledge of language assessment. On the teacher assessment knowledge test, the scores of the lecturers ($N = 152$) varied from 5 to 19, with a mean of 12.65 and a standard deviation (SD) of 3.788 as seen in Table 3 and Figure 1.

According to the descriptive statistics shown in Table 3, the assessment knowledge score data from 20 items and 152 Thai EFL university lecturers were normally distributed. The participants' mean score was 12.65, and the standard deviation was 3.788, indicating that there was variability in their test scores. Kurtosis and skewness values did not surpass the range of -2 to +2, and this was due to the fact that many participants performed well on the test. The skewness score of -.581 suggested that the distribution was slightly negatively skewed. The kurtosis score of -.836 suggested a relatively flat distribution. Figure 1 displays a histogram depicting the normal distribution of the 20-item test and 152 Thai EFL university lecturers. Consequently, the assessment test scores were appropriate for a norm-referenced decision.

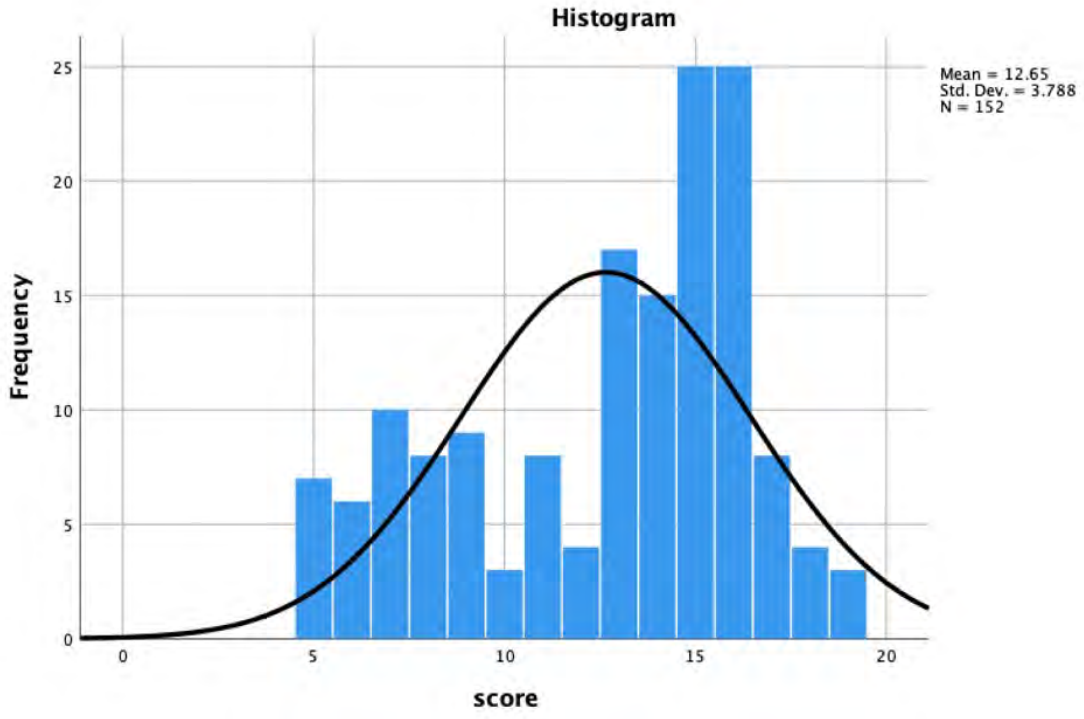
Table 3

Descriptive Statistics of the Assessment Knowledge Scores

N	M	SD	Range	Min	Max	SK	KU
152	12.65	3.788	14	5	19	-.581	-.836

Figure 1

Histogram Showing the Score Distribution of the 20-item Assessment Knowledge Test



To determine how lecturers did on the different test items, the mean number of correct responses per test item was computed and is shown in Table 4 below.

Table 4

Assessment Knowledge Test Items and Correct Answers

Items	Correct answer/Item	Mean score	Items	Correct answer/Item	Mean score
1	112	.737	**11**	120	.789
2	136	.895	**12**	96	.632
3	80	.526	**13**	104	.684
4	128	.842	**14**	56	.368
5	96	.632	**15**	40	.263
6	16	.105	**16**	17	.112
7	120	.789	**17**	120	.789
8	80	.526	**18**	104	.684

Items	Correct answer/Item	Mean score	Items	Correct answer/Item	Mean score
9	32	.211	**19**	128	.842
10	96	.632	**20**	128	.842

Table 4 reveals that the mean number of correct responses per test question ranged from 0.105 to 0.842. This can be interpreted in two different ways. First, lecturers did very well on items 4, 19, and 20, while their performance was lowest on item 6. Second, test items 4, 19, and 20 were the simplest, while question 6 was the most difficult.

Since the design of the assessment knowledge test was informed by Brookhart's (2011) teacher assessment skills framework and Eyal's (2012) work on digitally-mediated language assessment, it was deemed suitable to analyze the test results using these frameworks. Consequently, test items were analyzed in accordance with the corresponding statement in Brookhart's and Eyal's. Refer to Table 5 for a breakdown of the statements and related test item(s). When a statement was represented by many items, the average correct answer score was calculated. As a result, the statement's final score was determined by averaging the correct responses of lecturers to both questions.

Table 5

Lecturers' Test Scores according to Theoretical Frameworks

Themes (items)	Score/Item	Mean score/Item
1) understanding language learning and teaching (all)	90	0.592
2) articulating clear learning intentions (2)	136	0.895
3) communicating learning outcomes (2, 5)	116	0.763
4) using test methods, purposes, validity, fairness, and reliability (1, 3, 10) (12-14, 16-20)	95	0.625
5) analyzing test items (4,5)	112	0.737
6) communicating effective feedback (6)	16	0.105

Themes (items)	Score/Item	Mean score/Item
7) scoring (3, 15)	60	0.395
8) result-informed decision-making (7)	120	0.789
9) communicating scores with other stakeholders (8)	80	0.526
10) helping students plan their learning (9)	32	0.211
11) preparing tests (9)	32	0.211
12) digitally-mediated assessment literacy (11)	120	0.789

Table 5 reveals that participants were most literate in theme 2 (articulating clear learning intentions), themes 8 and 12 (result-informed decision-making and digitally-mediated assessment literacy), theme 3 (communicating learning outcomes), and theme 5 (analyzing test items), with average scores of .895, .789, .763, and .737 respectively. Moreover, they were moderately literate in theme 4 (using test methods, purposes, validity, fairness, and reliability), theme 1 (understanding language learning and teaching), and theme 9 (communicating scores with other stakeholders), with average scores of .625, .592, and .526, respectively. However, the findings also suggested that participants were the least literate in theme 7 (scoring), themes 10-11 (helping students plan their learning and preparing tests), and theme 6 (communicating effective feedback), with average scores of .395, .211, and .105, respectively.

4.2 Participants' Reported Assessment Beliefs

Closely connected to assessment knowledge are teacher beliefs about language assessment. It can be indicated that these beliefs are a critical component of assessment literacy. As seen in Table 6 below, there are nine aspects concerning beliefs towards their assessment.

Table 6

Participants' Reported Assessment Beliefs

Aspects	Mean	*SD*	Interpretation
Participants' beliefs towards their assessment			
Constructing and Administering Assessment	4.18	.407	Important
Performance Assessment	4.38	.413	Very Important
Scoring and Grading	4.11	.620	Important
Communicating Assessment Results with Others	4.00	.571	Important
Assessment Ethics	4.32	.563	Very Important
Digitally-mediated Assessment	4.11	.596	Important
Traditional Assessment Methods	3.55	.577	Important
Student-Involved Assessment	4.07	.755	Important
Non-achievement-based Grading	3.31	1.076	Undecided

According to Table 6, most participants found performance assessment and assessment ethics 'very important' when it came to language assessment in classrooms (M = 4.38 and 4.32, respectively). On the other hand, most of them believed that traditional assessment methods and non-achievement-based grading were least important among other aspects (M = 3.55 and 3.31, respectively).

4.3 Participants' Perceived Assessment Practices

While experience and knowledge influence beliefs, beliefs engender practice, which encompasses the entirety of language assessment literacy. Therefore, the last section of the survey examined participants' perceptions about assessment practices.

Table 7

Participants' Perceived Assessment Practices (from the most practiced order)

Aspects	Mean	*SD*	Interpretation
Informing students in advance how grades are to be assigned	4.53	0.697	used very often
Communicating classroom assessment results to students	4.53	0.612	used very often
Recognizing unethical, illegal, or otherwise inappropriate assessment methods	4.47	0.612	used very often
Developing assessments based on clearly defined course objectives	4.42	0.607	used very often
Communicating performance assessment criteria to students in advance	4.42	0.607	used very often
Matching assessments with classroom instruction	4.37	0.831	used very often
Recognizing unethical, illegal, or otherwise inappropriate uses of assessment information	4.37	0.684	used very often
Protecting students' confidentiality with regard to test scores	4.32	0.82	used very often
Weighing differently projects, exams, homework, etc., when assigning semester grades	4.32	0.82	used very often
Matching performance tasks to instruction and course objectives	4.26	0.562	used very often
Administering announced quizzes	4.21	0.713	used very often
Providing written feedback to students	4.16	0.765	used often
Choosing appropriate assessment methods for instructional decisions	4.11	0.567	used often
Revising previously produced teacher-made tests to match current instructional emphasis	4.11	0.809	used often
Using assessment results when evaluating class improvement	4.05	0.78	used often
Providing oral feedback to students	4.00	0.943	used often
Using assessment results when planning teaching	3.95	0.78	used often
Avoiding teaching to the test when preparing students for tests	3.84	0.898	used often

Aspects	Mean	*SD*	Interpretation
Following required procedures (time limit, no hints, no interpretation) when administering standardized tests.	3.79	1.084	used often
Incorporating attendance in the calculation of grades	3.79	1.228	used often
Assessing individual class participation	3.74	1.046	used often
Evaluating oral questions from students	3.63	0.895	used often
Using a table of specifications to plan assessments	3.58	1.121	used often
Providing traditional assessments (e.g., paper-pencil tests)	3.53	0.905	used often
Assessing students through observation	3.42	1.071	used often
Constructing a model answer for scoring essay questions	3.16	1.302	used occasionally
Interpreting standardized test scores (e.g., Percentile Rank) to students.	3.11	1.37	used occasionally
Administering unannounced quizzes	2.53	1.264	seldom used
Conducting item analysis (i.e., difficulty and discrimination indices) for teacher-made tests.	2.53	1.264	seldom used
Revising a test based on item analysis	2.47	1.219	seldom used

As seen in Table 7, participants reported that they very often informed their students in advance how grades were to be assigned ($M = 4.53$), communicated classroom assessment results to their students ($M = 4.53$), recognized unethical, illegal or inappropriate assessment methods ($M = 4.47$), develop assessments based on clearly defined course objectives ($M = 4.42$), and communicate performance assessment criteria to their students in advance ($M = 4.42$), respectively. However, surprisingly, they reported that they seldom conducted item analysis for teacher-made tests ($M = 2.53$) or revised a test based on item analysis ($M = 2.47$).

Table 8

Pair-wise comparisons for years of teaching and assessment practices

Groups (years)	Test Statistics	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
6-10 & 11-15	-4.000	8.640	-.463	.643	1.000
6-10 & 21 or more	-20.000	12.680	-1.577	.115	.688
6-10 & 1-5	59.000	10.024	5.886	.000	.000[a]
11-15 & 21 or more	-16.000	12.451	-1.285	.199	1.000
11-15 & 1-5	55.000	9.734	5.651	.000	.000[a]
21 or more & 1-5	39.000	13.449	2.900	.004	.022[a]

Note. [a] significant level is .05

To further analyze these findings, follow-up tests were done to examine pair-wise comparisons across the four teaching experience groups. For the assessment practices variable, significant differences were found 1) between the group with little teaching experience (six to ten years) and the group with the least teaching experience (one to five years), 2) between the group with some teaching experience (11 to 15 years) and the group with the least teaching experience (one to five years), and 3) between the group with the most teaching experience (21+ years) and the group with the least teaching experience (one to five years). This conclusion suggested that participants with more experience had more confidence in their use of assessment practices, while those with the least experience felt the contrary.

To examine the relationship between three variables, namely assessment knowledge, beliefs, and practices, 152 participants answered both language assessment knowledge test and questionnaires, and responses from both instruments were compared. To determine the relationship between teacher language assessment knowledge and their perceived assessment beliefs and practices, a correlation analysis was conducted.

Table 9

Correlation report between assessment knowledge test and practices

Descriptive Statistics	Mean	Std. Deviation	N
Test Score	12.65	3.788	152
Practices	3.8561	.43362	152

	Correlations	Test scores	practices
Test scores	Pearson Correlation	1	-.023
	Sig. (2-tailed)		.776
	N	152	152
Practices	Pearson Correlation	-.023	1
	Sig. (2-tailed)	.776	
	N	152	152

Note. **. Correlation is significant at the .01 level (2-tailed).

As seen in Table 9, a Pearson product-moment correlation coefficient was computed to assess the relationship between assessment knowledge test scores and assessment practices. There was a negative correlation between assessment knowledge test scores and assessment practices, [r = -.023, N = 152, p = .776]. This indicated that as participants' scores on the assessment knowledge test increased, their assessment practices tended to decrease. It is important to note that the term "negative correlation" implied an inverse relationship, where higher knowledge test scores were associated with lower assessment practices. This finding highlighted the need to further investigate the factors contributing to this negative relationship and to explore ways to bridge the gap between assessment knowledge and practical implementation in order to enhance the quality of assessment practices.

Table 10

Correlation report between assessment knowledge test and beliefs

Descriptive Statistics	Mean	Std. Deviation	N
Test Score	12.65	3.788	152
Beliefs	4.1826	.32746	152

Correlations		Test scores	Beliefs
Test scores	Pearson Correlation	1	.532**
	Sig. (2-tailed)		.000
	N	152	152
Beliefs	Pearson Correlation	.532**	1
	Sig. (2-tailed)	.000	
	N	152	152

Note. **. Correlation is significant at the .01 level (2-tailed).

As seen in Table 10, a Pearson product-moment correlation coefficient was computed to assess the relationship between assessment knowledge test scores and assessment beliefs. There was a significantly positive correlation between assessment knowledge test scores and assessment beliefs, [$r = .532$, $N = 152$, $p = .000$]. This suggested that participants with higher scores on the assessment knowledge test tended to hold more positive or favorable beliefs about assessment practices, while those with lower scores held fewer positive beliefs.

Table 11

Multivariate Analysis of Variance

	Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.996	16180.082[b]	2.000	136.000	.000	.996
	Wilks' Lambda	.004	16180.082[b]	2.000	136.000	.000	.996
	Hotelling's Trace	237.942	16180.082[b]	2.000	136.000	.000	.996
	Roy's Largest Root	237.942	16180.082[b]	2.000	136.000	.000	.996

	Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
score	Pillai's Trace	1.297	18.046	28.000	274.000	.000	.648
	Wilks' Lambda	.110	19.603[b]	28.000	272.000	.000	.669
	Hotelling's Trace	4.405	21.238	28.000	270.000	.000	.688
	Roy's Largest Root	3.274	32.035[c]	14.000	137.000	.000	.766

Note. a. Design: Intercept + score

b. Exact statistic

c. The statistic is an upper bound of F that yields a lower bound on the significant level.

Multivariate Analysis of Variance (MANOVA) in Table 11 was adopted to examine the differences in Thai EFL university lecturers' assessment knowledge in relation to their assessment beliefs and practices. The analysis revealed a significant difference in assessment beliefs and practices based on the assessment knowledge, $F(2,136) = 19.603$, $p = .000$; Wilk's lambda = .110, partial eta squared = .669. Furthermore, in Table 12, there is a significant effect of assessment knowledge on assessment beliefs, $F(14, 137) = 12.472$, $p = .000$, partial eta squared = .560. The results of the data analysis revealed a significant effect of assessment knowledge on participants' assessment beliefs. This suggested that participants' level of assessment knowledge significantly influenced their assessment beliefs. There was a significant effect of assessment knowledge on assessment practices, $F(14, 137) = 14.947$, $p = .000$, partial eta squared = .604. The data analysis revealed a significant effect of assessment knowledge on participants' assessment practices. This implied that participants' level of assessment knowledge had a substantial influence on their assessment practices.

Table 12

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Beliefs	9.073[a]	14	.648	12.472	.000	.560
	Practices	17.159[b]	14	1.226	14.947	.000	.604
Intercept	Beliefs	1692.764	1	1692.764	32576.850	.000	.996
	Practices	1479.948	1	1479.948	18049.185	.000	.992
score	Beliefs	9.073	14	.648	12.472	.000	.560
	Practices	17.159	14	1.226	14.947	.000	.604
Error	Beliefs	7.119	137	.052			
	Practices	11.233	137	.082			
Total	Beliefs	2675.342	152				
	Practices	2288.604	152				
Corrected Total	Beliefs	16.192	151				
	Practices	28.392	151				

Note. a. R Squared = .560 (Adjusted R Squared = .515)

b. R Squared = .604 (Adjusted R Squared = .564)

Overall, these findings highlighted the important role of assessment knowledge in shaping participants' beliefs and practices related to assessment, emphasizing the need for targeted interventions and professional development programs to enhance assessment knowledge among educators and promote more effective assessment practices.

Interestingly, if the MANOVA results were significant but the correlation was non-significant, it suggested that while there were differences between groups across multiple dependent variables (as indicated by MANOVA), there was no significant linear relationship between the specific variables being examined (as indicated by correlation).

5. Discussion

Apparently, language assessment knowledge plays a significant role for appropriate assessment practices in language classrooms (Popham, 2006).

Previous research has demonstrated that language instructors tend to have minimal or inadequate language assessment literacy (e.g., DeLuca & Johnson, 2017; Lam, 2015; Watmani et al., 2020; Yamtim & Wongwanich, 2014). Broadly speaking, this research seemed to resemble findings of past studies suggesting that language instructors had low levels of language assessment literacy. Before embarking on the discussion, it is worth noting that this study adapted a measure called teacher assessment knowledge test, designed specifically to measure language instructors' language assessment knowledge based on Brookhart's (2011) framework and Eyal's (2021) work. Looking back, the scores of EFL university lecturers ($N = 152$) ranged from 5 to 19 with an average of 12.65 and a standard deviation (SD) of 3.788. Despite the fact that the majority of participants considered assessment to be vital, their language assessment knowledge remained weak in many assessment domains (see Table 2). According to the previously presented findings, participants were highly literate and skilled in articulating clear learning intentions, although this result seemed to contradict what was neglected in the 1990 standards. This could be supported by Edwards (2013), who demonstrates that teachers must communicate their intents and aims to their students as the first step in a successful assessment plan (Kennedy, 2008). Since learning intentions are closely related to the focus of what is taught and assessed, establishing clear learning intentions helps pave the way for students to master key concepts and learning progressions.

Although studies concerning language assessment literacy are limited, this study seemed to contradict the findings of Davidheiser (2013) indicating that teachers have limited knowledge of how technology is involved in assessment. Moreover, the findings from this study revealed that university lecturers believed that digitally-mediated language assessment was important for them. The explanation for this contradiction may be that participants in this study may have experienced the use of technology for assessment in virtual classrooms due to the pandemic, thus this might have allowed them to learn more and better prepare for

the utilization of technology in their assessment. Despite the fact that the use of technology in general does not often indicate effective technology use in assessment, the findings of the current research raised significant issues concerning technology use in assessment practices. First, a constraint might be a matter of methodology. This study's instrument was based on an adaptation of Eyal's (2012) index of teacher digital assessment literacy, which included three levels: basic, intermediate, and advanced. The items were matched to Eyal's levels in accordance with the digital affordances of the study's context. Perhaps the number of questions was insufficient to represent participants' real assessment skills and classroom practices.

Another significant finding was that participants were *the least literate* in communicating effective feedback. This may be attributed to constraints, such as time constraints, that impacted and limited their practices. Simply put, these lecturers may not have had sufficient time to provide effective feedback to their students. This finding was consistent with the finding reported by Wiboolyasarin's (2021) that Thai EFL instructors stated that time constraints hindered them from providing feedback to students and that they had little opportunities to do so, which is also in congruence with the findings of Mao and Crosthwaite (2019) and Lee (2019). In addition, Thirakunkovit (2019) has reported that class size was a significant factor in providing effective feedback, and the majority of his participants believed that providing feedback to their students had a negligible effect on their learning, as students typically made the same errors on subsequent assignments. Significantly, his research revealed a dearth of substantial knowledge and limited assessment pedagogical practices about the delivery of effective feedback.

Looking at the beliefs, it is evident that participants in this research considered that performance assessment was quite significant for classroom assessment. This seemed to be consistent with the results from the assessment

practice section, which indicated that participants often reported employing performance assessment and conveying performance assessment criteria in advance to students. It is generally believed that when teachers convey assessment criteria in advance to the students, it can be implied that they are aware of classroom assessment practices and its principles (Airasian, 2000). When students are informed of assessment criteria in advance, they are also aware of what they will be assessed on, and thus able to improve themselves and prepare for it. Chinda (2014) provides evidence that Thai EFL instructors saw performance assessment as advantageous for their students. However, these results contradicted Watson Todd's (2019) research, indicating that Thai EFL instructors continue to rely more on test-style assessments (final exam, midterm examinations, quizzes) than continuous assessment (performance/alternative assessment). However, research has proposed investigating and delving further into the reasons why the majority of language instructors choose these methods in order to obtain a deeper knowledge of their existing practices (Imsa-ard, 2023). Moreover, according to Chinda (2014), his participants did not believe that performance-based assessment could be performed successfully owing to class size, and several participants reported lacking appropriate assessment knowledge.

Interestingly, the majority of participants found non-achievement-based grading to be the least common in this study. To elaborate, non-achievement-based grading includes aspects such as student effort in the grading. This study's result was not surprising but puzzling. More than half of the participants in this research used non-achievement-based grading often to very often. This was consistent with the results of prior research (Cheng & Sun, 2015; Davidheiser, 2013). What is puzzling is that it is advantageous to discuss the inclusion/exclusion of non-achievement-based grading in their assessment, given that non-achievement-based grading might result in construct-irrelevant variances and so make grade interpretations quite deceptive (Cheng & Sun, 2015).

6. Conclusion, Limitations, and Implications

This study examined the language assessment knowledge, beliefs, and practices of Thai EFL lecturers. In contrast to the lack of assessment knowledge reported by scholars (e.g., Popham, 2009; Stiggins, 2002), our study results presented a little more optimistic picture about the knowledge of language assessment among EFL lecturers. Despite deficiencies in some assessment domains, the majority of participants performed rather well at the moderate level on the test of language assessment knowledge. The results indicated that lecturers were most competent at articulating clear learning intentions, making result-informed decisions, digitally-mediated assessment literacy, communicating learning outcomes, and analyzing test items, in that order. However, they were the least competent at scoring, assisting students with learning planning, preparing tests, and providing effective feedback. Regarding beliefs, the majority of lecturers stated that performance assessment and assessment ethics were crucial in language assessment. These beliefs were also mirrored in their perceived assessment practices, in which they reported applying performance assessment often and providing students with performance assessment criteria beforehand.

Since the purpose of our study was to examine the language assessment literacy of university lecturers in a Thai context, we gained evidence and insight that will allow us to pursue this topic further in future qualitative research. To support lecturers, it is important to establish some professional development related to assessment practice that are practical for them. However, there are a few limitations that should be considered. First, caution is warranted when interpreting the findings, since we relied only on self-reported data from lecturers, who may have provided a more favorable view of their assessment skills out of social desirability. Second, our research lacked a comprehensive and representative sample of Thai EFL university lecturers, which limited the generalizability of the findings. Moreover, the responses to our study's research questions generated other study-worthy questions. Case-study research with

sufficient detail to derive context-specific findings and differentiate linkages between individual factors, personal experiences, and institutional and social contexts may be more suited to answering such concerns. Our intention and aim are that this research will serve as a reference point for future in-depth investigations on language assessment literacy.

7. About 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. He can be reached at 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). He can be reached at supong.t@litu.tu.ac.th

8. Acknowledgement

I am sincerely appreciative of the constructive feedback provided by anonymous reviewers as well as the editorial team of the PASAA journal. Also, I would like to express my heartfelt appreciation to all of my PhD committee members for their useful comments to help improve the quality of this work. Any opinions, findings, and conclusions or recommendations demonstrated in this article are those of the authors and do not reflect the views of the university.

9. References

Acar–Erdol, T., & Yildizli, H. (2018). Classroom assessment practices of teachers in Turkey. *International Journal of Instruction, 11*(3), 587–602.

Airasian, P. W. (2000). *Assessment in the classroom: A concise approach*. McGraw–Hill Company.

Al–Bahlani, S. M. (2019). *Assessment literacy: A study of EFL teachers' assessment knowledge, perspectives, and classroom behaviors* [Doctoral dissertation, The University of Arizona, Arizona, USA].

Alghammas, A. (2020). Online language assessment during the COVID–19 pandemic: University faculty members' perceptions and practices. *Asian EFL journal*, *27*(44), 169–195.

Alkharusi, H. (2009). Classroom assessment environment, self–efficacy, and mastery goal orientation: A causal model. *INTI Journal: Special Issue on Teaching and Learning,* 104–116.

Alkharusi, H. (2011). A logistic regression model predicting assessment literacy among in–service teachers. *Journal of Theory and Practice in Education, 7*(2), 280–291.

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://dx.doi.org/10.5539/jel.v1n2p217

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

Bachman, L. F. (2002). Some reflections on task–based language performance assessment. *Language Testing, 19*(4), 453–476. https://doi.org/10.1191/0265532202lt240oa

Barnes, N., Fives, H., & Dacey, C. M. (2014). Teachers' beliefs about assessment. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 296–312). Routledge.

Bashitialshaaer, R., Alhendawi, M., & Avery, H. (2021). Obstacles to applying electronic exams amidst the COVID–19 pandemic: An exploratory study in the Palestinian universities in Gaza. *Information, 12*(6), 256. https://doi.org/10.3390/info12060256

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

Birkett, M. A., & Day, S. J. (1994). Internal pilot studies for estimating sample size. *Statistics in medicine*, *13*, 2455–2463. https://doi.org/10.1002/sim.4780132309

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

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, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement–oriented resources heighten conceptions of assessment as school accountability. *Journal of MultiDisciplinary Evaluation, 6*(12), 68–91. https://doi.org/10.56645/jmde.v6i12.236

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

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

Cheng, L., Rogers, W. T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education, 33*(1), 9–32. http://dx.doi.org/10.1080/02602930601122555

Cheng, L., & Sun, Y. (2015). Teachers' grading decision making: Multiple influencing factors and methods. *Language Assessment Quarterly*, *12*(2), 213-233. https://doi.org/10.1080/15434303.2015.1010726

Chinda, B. (2014). Teachers' reactions towards performance–based language assessment. *PASAA, 48*, 57–88.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Connelly, L. M. (2008). Pilot studies. *Medsurg nursing, 17*(6), 411–412.

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn?. *Language Testing in Asia, 10*, 1–16. https://doi.org/10.1186/s40468–020–00101–6

Csépes, I. (2021). The evolving concept of (language) assessment literacy. Implications for teacher education. *Central European Journal of Educational Research, 3*(1), 120–130. https://doi.org/10.37441/CEJER/2021/3/1/9360

Davidheiser, S. A. (2013). *Identifying areas for high school teacher development: A study of assessment literacy in the Central Bucks School District* (Doctoral dissertation, Drexel University).

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

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".

Edwards, F. (2013). Quality assessment by science teachers: Five focus areas. *Science Education International*, *24*(2), 212–226.

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

Ha, X. V., & Murray, J. C. (2021). The impact of a professional development program on EFL teachers' beliefs about corrective feedback. *System, 96*(102405), 1–14. https://doi.org/10.1016/j.system.2020.102405

Ha, X. V., Tran, N. G., & Tran, N. H. (2021). Teachers' beliefs and lractices regarding assessment in English as a foreign language classrooms in Vietnam. *The Qualitative Report, 26*(11), 3457–3475. https://doi.org/10.46743/2160–3715/2021.5063.

Hussain, S., Shaheen, N., Ahmad, N., & Islam, S. U. (2019). Teachers' classroom assessment practices: challenges and opportunities to classroom teachers in Pakistan. *Dialogue, 14*(1), 87–97.

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. (2021). Choices of teaching practices: Does the English national examination cause Thai EFL teachers to teach to the test? *TESOL International Journal, 16*(5), 56–86.

Imsa–ard, P. (2023). Digitally-mediated Language Assessment Practice (D-LAP): Qualitative Case Studies of Four Thai EFL University Lecturers. *REFLections, 30*(2), 403–427. https://so05.tci-thaijo.org/index.php/reflections/article/view/267222

Inbar–Lourie, O., & Donitsa–Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language.

Assessment in Education: Principles, Policy & Practice, 16(2), 185–204. http://dx.doi.org/10.1080/09695940903075958

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. http://dx.doi.org/10.18823/asiatefl.2020.17.2.27.689

Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non–language testers?. *Language Testing*, *30*(3), 345–362. https://doi.org/10.1177/0265532213480334

Julious, S. A. (2005). Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, *4*(4), 287–291. https://doi.org/10.1002/pst.185

Kang, H. (2021). Sample size determination and power analysis using the G* Power software. *Journal of educational evaluation for health professions*, *18*, 1–12. https://doi.org/10.3352/jeehp.2021.18.17

Kaur, A., Young, D., & Kirkpatrick, R. (2016). English education policy in Thailand: Why the poor results?. In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 345–361). Springer.

Kennedy, D. (2008). Linking learning outcomes and assessment of learning of student science teachers. *Science Education International*, *19*(4), 387–397.

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

Lee, I. (2019). Teacher written corrective feedback: Less is more. *Language Teaching*, *52*(4), 524–536. https://doi.org/10.1017/S0261444819000247

Levi, T., & Inbar–Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly, 17*(2), 168–182. https://doi.org/10.1080/15434303.2019.1692347

Mao, S. S., & Crosthwaite, P. (2019). Investigating written corrective feedback: (Mis)alignment of teachers' beliefs and practice. *Journal of Second Language Writing*, *45*, 46–60. https://doi.org/10.1016/j.jslw.2019.05.004

Narathakoon, A., Sapsirin, S., & Subphadoongchone, P. (2020). Beliefs and classroom assessment practices of English teachers in primary schools in Thailand. *International Journal of Instruction, 13*(3), 137–156.

Newby, P. (2014). *Research methods for education*. Routledge.

Noijons, J. (2013). Testing computer assisted language testing: Towards a checklist for CALT. *CALICO Journal, 12*(1), 37-58. https://doi.org/10.1558/cj.v12i1.37-58

Norris, J. M. (2016). Current uses for task–based language assessment. *Annual Review of Applied Linguistics, 36*, 230–247. http://dx.doi.org/10.1017/S0267190516000027

O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing, 30*(3), 363–380. https://doi.org/10.1177/0265532213480336

Pallant, J. (2005). *SPSS survival guide: A step by step guide to data analysis using SPSS for windows* (3rd ed.). Open University Press.

Popham, J. W. (2006). Needed: A dose of assessment literacy. *Educational leadership, 63*(6), 84–85.

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

Qian, D. D. (2008). English language assessment in Hong Kong: A survey of practices, developments and issues. *Language Testing, 25*(1), 85–110. https://doi.org/10.1177/0265532207083746

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). Preparing teachers to use the enabling power of assessment. In C. Wyatt–Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 303–323). Springer.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan, 72*(7), 534–539.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 3*(77), 238- 245.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational measurement: Issues and practice, 18*(1), 23–27. https://doi.org/10.1111/j.1745–3992.1999.tb00004.x

Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758-765.

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

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.

Thirakunkovit, S. (2019). Exploring writing assessment literacy and classroom practices of Thai University instructors. *Journal of AsiaTEFL*, *16*(4), 1135–1151. http://dx.doi.org/10.18823/asiatefl.2019.16.4.5.1135

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 127–146). Macmillan Publishing.

Thong–Iam, N., & Subphadoongchone, P. (2019). Professional development needs of Thai EFL university teachers: A focus on the assessment literacy and practice of student self–assessment in classrooms. *MANUTSAT PARITAT: Journal of Humanities*, *40*(2), 56–72.

Viengsang, R. (2016). Exploring pre-service English teachers' language assessment literacy. *Modern Journal of Language Teaching Methods, 6*(5), 432-455.

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

Wang, T. H., Wang, K. H., & Huang, S. C. (2008). Designing a web–based assessment environment for improving pre–service teacher assessment literacy. *Computers & Education, 51*(1), 448–462. http://dx.doi.org/10.1016/j.compedu.2007.06.010

Wang, Y. H. (2014). Use of interactive web–based exercises for English as a foreign language learning: Learners' perceptions. *Teaching English with technology*, *14*(3), 16–29.

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

Watson Todd, R. (2019). How is English assessed at Thai schools?. *THAITESOL Journal, 32*(1), 1–15.

Wiboolyasarin, W. (2021). Written corrective feedback beliefs and practices in Thai as a foreign language context: A perspective from experienced

teachers. *Language Related Research, 12*(3), 81–119. https://doi.org/10.29252/LRR.12.3.4

Wiratchai, N. (2012). Interested sample size and statistical analyzes determination. *Journal of Sport and Health Science, 5*(3), 315–321.

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. http://dx.doi.org/10.18823/asiatefl.2019.16.2.13.653

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. http://dx.doi.org/10.58379/UZON5145

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, *43*(3), 492-513. https://doi.org/10.1002/j.1545-7249.2009.tb00246.x

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

Zhang, Z., & Burry–Stock, J. A. (1994). *Assessment practices inventory*. The University of Alabama.

Zhang, Z., & Burry–Stock, J. A. (2003). Classroom assessment practices and teachers' self–perceived assessment skills. *Applied measurement in education*, *16*(4), 323–342. https://doi.org/10.1207/S15324818AME1604_4