

Content list available at <http://ijltr.urmia.ac.ir>

*Iranian Journal
of
Language Teaching Research*
ORIGINAL ARTICLE



Urmia University

Assessment Literacy Among Indonesian Pre-service English Language Teachers: A Mixed-Methods Study

Andrzej Cirocki ^{a,b}, Syafi'ul Anam ^{b,*}, Nur Arifah Drajadi ^c, Bill Soden ^d

^a *University of York, United Kingdom*

^b *Universitas Negeri Surabaya, Indonesia*

^c *Universitas Sebelas Maret, Indonesia*

^d *Independent Scholar, United Kingdom*

ABSTRACT

Using a sequential explanatory design, this study investigated levels of assessment literacy among pre-service English as a foreign language teachers in Indonesia and their views on how their undergraduate teacher education programmes prepared them for classroom assessment. A total of 320 pre-service teachers completed a Language Assessment Literacy (LAL) questionnaire, and 37 participated in semi-structured interviews. Findings indicated moderate overall LAL, with significant differences based on university status and affiliation, favouring state and Ministry of Education-affiliated universities. There was general satisfaction with education programmes but a desire for more practical training, especially in private and Ministry of Religious Affairs-affiliated universities. Recommendations included increasing practical assessment training, incorporating alternative methods, and integrating technology and Artificial Intelligence into the curriculum. These findings highlight the need to enhance and standardise LAL training to prepare pre-service teachers for effective classroom assessment in Indonesia.

Keywords: assessment literacy; assessment competence; pre-service teachers; English language teaching; Indonesia

© Urmia University Press

ARTICLE HISTORY

Received: 3 Nov. 2024

Revised version received: 25 Jan. 2025

Accepted: 10 Feb. 2025

Available online: 10 Mar. 2025

* Corresponding author: Department of English, Faculty of Languages and Arts, Universitas Negeri Surabaya, Indonesia

Email address: syafiul.anam@unesa.ac.in

© Urmia University Press

10.30466/ijltr.2025.55691.2851

Introduction

Assessment competence has recently become a key indicator of teacher professionalism (Bachman & Damböck, 2018; Cirocki & Hallet, 2024; Kremmel & Harding, 2019; Papageorgiou & Bailey, 2019; Yeo, 2021). It is a complex construct that typically encompasses teachers' (1) knowledge and understanding of assessment principles, procedures, and expectations; (2) conduct of assessment in terms of teaching, implementation (including student engagement), and evaluation; (3) understanding and utilising the information gathered through assessment practices; and (4) awareness of the impact and challenges of assessment (Coombe et al., 2020; Hay & Penney, 2013; Kremmel & Harding, 2019; Pastore & Andrade, 2019). Put simply, language teacher assessment literacy (AL) entails comprehension, application, and interpretation of and critical engagement with assessment (Hay & Penney, 2013).

This complexity challenges both English as a foreign language (EFL) teachers and teacher educators who prepare pre-service teachers for schoolwork. In Indonesia, the context of the current study, language assessment is gaining importance in educational discourse and has recently appeared in policies and government-issued documents (Ministry of Education, Culture, Research, and Technology, 2022). Pre-service EFL teachers are required to complete assessment-oriented courses as part of their teacher education programmes (Directorate of Teacher Profession Education, 2024). The primary purpose of these courses is to equip future teachers with the knowledge and skills to conduct assessments in primary and secondary schools. Although these courses vary from university to university, they focus on basic aspects of classroom assessment and lessen the importance of exam-oriented education. In the future, more consistency, uniformity, and detailed content within assessment courses across the country will be extremely useful. Similarly, more systematic research on assessing students' English language competence at schools and EFL language teachers' AL is vital. Existing studies often lack rigorous design or reporting or are published locally rather than internationally (e.g., Nyudak et al., 2022; Putra et al., 2024; Suherman, 2022). Hence, this mixed-methods study fills this gap.

Accordingly, this article focuses on language assessment literacy (LAL) and its importance in twenty-first-century EFL classrooms and EFL teacher education. It begins by defining the concept of AL and discussing its origins. It also synthesises previous research on teacher AL. Following this, the article presents a mixed-methods study conducted among Indonesian pre-service EFL teachers. The purpose of this study was threefold: (1) to measure levels of LAL among pre-service teachers, identifying variations due to university status and affiliation; (2) to ascertain the extent to which undergraduate teacher education programmes have prepared them for successful implementation of classroom assessment; and (3) to elicit recommendations from pre-service teachers regarding improving the integration of LAL into their teacher education programmes. The article ends with implications for EFL teacher educators and policymakers.

Literature Review

Evolution and Conceptualisation of LAL: A Brief Overview

This section briefly introduces the development of AL in general education, particularly in view of notions of summative and formative assessment. It then provides a more in-depth exploration of LAL's componential and development conceptualisation, with a review of research into teacher LAL and its implications for professional development. Whilst quantitative methods tend to dominate, mixed-methods and qualitative approaches have also been applied with some success.

In charting the evolving conceptualisation of both AL and LAL, it is useful to bear in mind the two key concepts within assessment theory, summative and formative assessment. Summative assessment, often referred to as assessment of learning, relates to traditional, often high-stakes tests where test scores are paramount and used to arrive at go/no-go decisions (Popham, 2009). By contrast, formative assessment, often known as assessment for learning (Black & Wiliam, 1998), refers to ongoing instructional activities with an improvement orientation. It comprises classroom-based assessment and feedback aimed at helping students improve learning and outcomes and enhancing teachers' own content and practice.

Stiggins (1999) has often been cited as the first to embrace AL, as he outlined the importance of a range of teacher assessment practices related mainly to the psychometric testing approaches of that time, emphasising summative assessment purposes. Over the following decades, AL has become increasingly understood as a differentiated and situated professional competence, with research focusing mainly on teachers and using quantitative survey methods (e.g., DeLuca & Klinger, 2010; Mertler & Campbell, 2005). A significant contribution by Popham (2009) provided a conceptualisation of AL that proposed dividing AL into two main areas, classroom assessment and accountability assessments, thereby fully embracing both the summative and formative aspects of AL. At this point, LAL itself began to emerge as a separate concept (DeLuca et al., 2019).

At the end of the first decade of this century, research into LAL could be described as being "in its infancy" (Fulcher, 2012, p. 117); however, it has attracted much more attention in recent years. In an early and influential attempt to define and develop the concept of LAL, Davies (2008) focused predominantly on language tests and a skills/knowledge approach. However, he also stressed the need to take account of principles of assessment or fairness, and the appropriate use of test results, with an understanding of the ethical issues they raise. In this regard, the notion of skills refers to training in test writing and analysis, including statistical measurement, with knowledge comprising an understanding of language learning and teaching within the relevant assessment context.

Around the same time as Davies' (2008) study, a key paper by Inbar-Lourie (2008) articulated a significant move away from the predominantly test and summative assessment focus of LAL to construct a view of assessment as a more wide-ranging concept, including alternatives (e.g., classroom assessments such as portfolios or oral presentations) and greater emphasis on assessment for learning processes and the social context of assessment. In the following year, Davison and Leung's (2012) review of teacher-based assessment (TBA) focused principally on teacher-led assessment for learning. Davison and Leung (2012) sought to problematise formative assessment by highlighting several misconceptions surrounding it such as assumptions that any classroom assessment is formative or that alternative forms of assessment, such as oral presentations, are part of externally set summative examinations. They pointed to how TBA "calls into action a multifaceted combination of linguistic, pragmatic, and cultural resources" (Davison & Leung, 2012, p. 406). Davison and Leung's (2012) review called for a better theorisation of TBA, with the need to deal with the challenge of language as a content carrier and the implications of how to assess such content and the variable nature of TBA, which is highly context-specific and context-dependent.

The studies reviewed above reveal a developing conceptualisation of LAL, but also point to a lack of agreement on its fundamental components. In view of this, Fulcher (2012), in his study on the assessment training needs of language teachers, argued for a wider conception of LAL by combining three key aspects: measurement knowledge, assessment know-how, and the social perspective in assessment and testing. Fulcher (2012) emphasised the importance of relating skills and knowledge of processes and principles to the wider social and historical context to understand where practices originate and to critically evaluate their impact on society, institutions,

and individuals. He also argued for an integrated approach to LAL that considers knowledge and skills related to both summative and formative assessment processes.

Fulcher's (2012) contribution was taken up quickly with a special edition on LAL in the *Language Testing Journal* in 2013. Taylor's (2013) review article in this edition pointed to how earlier definitions of LAL comprising broad sets of components and competences did not indicate the nature of expertise or depth of knowledge needed by different groups of stakeholders. Taylor (2013) welcomed Pill and Harding's (2013) contribution in the same journal issue, which identified a nominal literacy level of a particular group of stakeholders regarding assessment processes and tests for doctors in Australia. Pill and Harding (2013) provided a breakdown of a continuum from illiteracy through functional and procedural literacy to multidimensional literacy but left the nature and content of these levels rather vague (Tsagari, 2020).

Taylor (2013) also argued for a view of LAL as an integrated concept, supporting Fulcher (2012) and Inbar-Lourie (2013). Inbar-Lourie (2013) raised concerns that some conceptions of LAL understated or ignored the language trait and the pedagogical elements required by teachers, arguing instead for creating a special knowledge base that merged general assessment knowledge with language-related expertise. Later, Inbar-Lourie (2016) introduced a multidimensional view of LAL encapsulated in the term "Language Assessment Literacies", a looser and more dynamic description that better reflects the local contexts in which LAL operates (Tsagari, 2020).

Taylor's (2013) main contribution, however, was her concentric circles model designed to consider the various stakeholders in LAL and the nature and degree of expertise in the competences they require (Figure 1). In doing so, she provided a means to profile stakeholders and facilitate a clearer "focus on pedagogical efforts" (Tsagari, 2020, p. 17). It should be noted that, in this model, language pedagogy is prioritised over technical skills and principles.

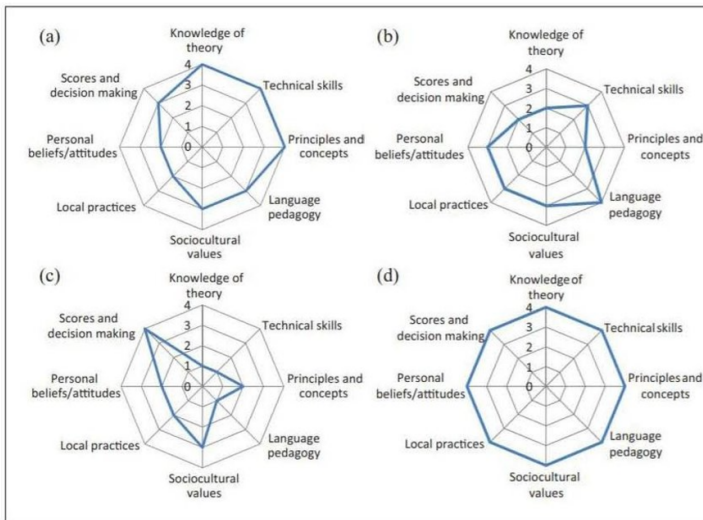


Figure 1. Concentric Circles Model of LAL for Teachers (Taylor, 2013, p. 410)

Note. LAL profiles for four different types of stakeholders: (a) profile for test writers, (b) profile for classroom teachers, (c) profile for university administrators, and (d) profile for professional language testers.

In summary, existing scholarship suggests an understanding of LAL as multicomponential with a developmental continuum differentiated according to stakeholders' needs (Kremmel & Harding, 2019; Taylor, 2013; Tsagari, 2020). This implies that an assessment-literate English teacher will be able to determine workable aims for ethically sound assessment, reflecting an understanding of assessment policies in their own local contexts and washback in relation to their own teaching and assessment practices. Additionally, they will possess the know-how to design, administer, and analyse assessments, and to reflect on their beliefs in relation to these, as well as the ability to interpret results in relation to student achievement and learning, along with their own pedagogical methods and instruction (Fulcher, 2020). For more information on these elements and factors, see Kremmel and Harding (2019).

Previous Research on Teacher LAL: Levels and Training Needs

Research into teacher LAL tends to be divided around two interrelated foci: the state of specific groups of teachers' LAL knowledge and expertise, and a more recent and growing interest in exploring pre-service teachers' LAL, along with the training provision for teachers (Fulcher, 2020; Giraldo, 2023).

Among studies on in-service English teachers conducted in various countries and contexts, there seems to be a tendency to equate assessment and evaluation with testing to measure student learning, along with a lack of awareness of both assessment for learning options in practice and the concept of LAL itself (Sevimel-Sahin, 2020). One example is provided by Kioms et al.'s (2011) study of Iranian EFL teachers, which found that low levels of LAL correlated closely with washback effects and that, despite feeling prepared for teaching and assessment, major misunderstandings existed regarding the assessment. In another Iranian context, Tajeddin et al. (2018) reported that novice and experienced teachers shared some aspects of speaking assessment literacy, with experienced teachers demonstrating more consistency in their assessment practice. In the British context, Berry et al. (2017) conducted interviews and observations with EFL teachers, followed by larger focus group discussions to assess factors contributing to LAL effectiveness and perceived needs. Based on Davies' (2008) conception of LAL, Berry and colleagues (2017) found that teachers did not engage with the term "assessment", preferring instead to use the term "testing", perhaps reflecting a perception of testing as imposed top-down. They also found that experience was the key to developing assessment practices, not training, and further concluded that their participants were unfamiliar with the term AL, revealing a gap between testing experts/researchers and teachers. In one of only a few studies conducted in the Indonesian context, Zulaiha et al. (2020) reported a moderately highly professed level of LAL among EFL teachers. Notably, interviews and documentary evidence highlighted the impact of local policies with more of a focus on grades over the use of multiple-choice tests and some doubt over the quality of assessments offered.

Xu and Brown (2016) used a survey instrument to explore LAL levels among Chinese English language university teachers and reported a basic level of LAL with no clear links to contextual factors, a finding echoed by Puspawati's (2019) mixed-methods study of Indonesian university teachers. The latter study revealed higher confidence in principles than in assessment skills. Contrasting slightly with these findings, a later mixed-methods study with middle-school EFL teachers by Fitriyah et al. (2022), in the same Indonesian context but focusing on classroom-based LAL, reported a level of confidence in technical and pedagogical skills, but less confidence in principles and theory. In the Turkish context, but similarly focusing on university teachers, Mede and Atay (2017) found not only a low level of reported LAL, but also a surprising lack of knowledge of the concepts of validity, reliability, and confidence in relation to traditional

grammar and vocabulary testing. They also highlighted the need for more training on formative classroom-based assessment practices. A qualitative study by Giraldo (2019) highlighted the complex and multifaceted nature of the LAL concept among Colombian teachers in a university institute but reported more positive findings which pointed to a degree of LAL with teachers embracing both summative and formative assessment practices, revealing the importance of past experiences in forming their beliefs around LAL.

Some studies have highlighted technological challenges along with challenges to teachers arising from the social, cultural, and political context in which they teach (Inbar-Lourie & Donitsa-Schmidt, 2009; Shim, 2009), whilst others have identified issues with education systems and top-down control depriving teachers of autonomy (Rezagah, 2022). For example, Shim's (2009) doctoral study of Korean primary EFL teachers revealed a level of conviction around LAL principles, but a professed inability to put these into practice due to factors beyond teachers' control such as overcrowded classrooms, bureaucracy, lack of funding for language teaching, and heavy workloads.

In relation to training needs and provision, Weng and Shen's (2022) study highlighted the lack of assessment content in training courses for pre-service teachers and the generally insufficient preparation for assessment for in-service teachers across multiple contexts. These observations are borne out by evidence from earlier studies. For instance, a large-scale study of EFL teachers across seven European countries conducted by Vogt and Tsagari (2014) revealed low levels of AL and a perceived need for training. Their interview data highlighted dissatisfaction with pre-service and in-service training and participants' inability to critically evaluate tests, using more traditional forms of assessment at the expense of informal formative practices. A further key finding illuminated the way that reliance on "on-the-job" development of knowledge and skills could lead to the risk of "testing as you were tested" (Vogt & Tsagari, 2014, p. 391). In a later study, Tsagari and Vogt's (2017) data from secondary EFL teachers in Greece, Cyprus, and Germany revealed low reported levels of LAL and a similar reliance on traditional assessment procedures. Notably, Tsagari and Vogt (2017) found that most of the teachers could not identify training and development needs in LAL and reported having learnt little or nothing from pre-service training. By contrast, a recent Indonesian study on EFL middle-school teachers (Fitriyah et al., 2022) identified a functional self-reported level of classroom-based LAL, but with a reported need for professional training in theory and assessment principles, regardless of whether teachers were experienced or novices. Strikingly, in a Colombian study by Giraldo (2019), five teacher participants identified needs which were somewhat differentiated according to their assessment experiences.

Kremmel and Harding's (2019) study of self-perceived needs for the successful application of LAL largely supported Taylor's (2013) model in that teacher self-reported perceptions generally matched Taylor's (2013) predictions for a teacher LAL profile. Understanding assessment policy, local practices, and socio-cultural values were lower than her model predicted but, conversely, teachers rated knowledge of language theory and understanding testing principles as more important than the model profile predicted. In China, an interview-based study by Yan et al. (2018) found stronger development and training needs in assessment practice than in theory (see also Berry et al., 2017, above). Their study once again confirmed to some extent Taylor's (2013) LAL teacher profile proposal, demonstrating how that profile is mediated by the interplay between contextual and experiential factors.

Interest in teacher demographics or differences between in-service and pre-service teachers has been extremely recent (Tsagari, 2020). Several studies have focused on the impact of training programmes on teacher LAL (Giraldo, 2021; Giraldo & Murcia, 2018; Hildén & Fröjdendahl, 2018; Levi & Inbar-Lourie, 2019). For example, Levi and Inbar-Lourie (2019) highlighted how the partial application of generic principles and skills was accompanied by a lack of attention given to

language-related constructs. In a small-scale study with pre-service Finnish language teachers following a unit of instruction on assessment, Hildén and Fröjndendahl (2018) reported a boost in self-esteem and greater confidence to assess aspects of language learning with a learner-centred focus. Giraldo and Murcia (2018) conducted an action research study on the impact of a language assessment course in Colombia, focusing on what pre-service teachers need and want from such courses. They found that there was a demand for practical assessment skills but no interest in principles and ethics. Giraldo (2021) later reviewed 14 studies on training programmes, concluding that whilst many training attempts tended to improve teachers' awareness of the need to assess all skills, language principles were largely ignored. Inbar-Lourie (2019) also reached this conclusion in a review of other studies on teacher involvement in language testing.

The findings from this research on actual training programmes highlight the ongoing need for teacher educators to create viable integrated LAL training programmes with a methodology in practice. Fulcher (2020), for example, argued for an apprenticeship approach to teacher training grounded in theory, which would enable testing and assessment to be conducted in local contexts via test development cycles, thereby introducing key principles and theory in context (Harding & Kremmel, 2016; McMillan, 2011).

Summary of the Literature Review

In summary, there has been agreement around several elements in LAL such as skills, knowledge, principles, and awareness of the wider social and historical context in which assessment occurs (Davies, 2008; Fulcher, 2012; Taylor, 2013; Tsagari, 2020). LAL has been conceptualised as a continuum, with degrees of literacy possible (Pill & Harding, 2013; Taylor, 2013; Tsagari, 2020), whereas differentiation of content and focus for different stakeholder groups is required, with a particular gap in research on pre-service teachers' LAL, especially in the Indonesian context. Both summative and formative approaches are agreed to be central and understood within local teaching and assessment contexts. The linguistic element in terms of theoretical knowledge and practical pedagogical knowledge has been recognised as a key component in LAL, but with the changing conceptualisation of the language construct over time, there has been no agreement to date as to what this constitutes (Harding & Kremmel, 2016; Inbar-Lourie, 2016; Tsagari, 2020).

Research findings on self-reported beliefs and perceptions suggest that a low or basic level of LAL is not uncommon in teachers in multiple countries, with studies in the Indonesian context revealing, at best, a self-reported functional level of LAL. Teachers are more confident in traditional summative and high-stakes testing than classroom-based formative assessments. Research has frequently pointed to the need for more focus on both pre-service and in-service training programmes, and although teachers generally show a preference for a practical focus, in the Indonesian context, a need for more theoretical elements has also been expressed (Fitriyah, 2022). In view of the above, the research questions for the current study were as follows:

1. What are the existing levels of LAL among Indonesian pre-service EFL teachers, and do these levels vary according to university status (private/state) and university affiliation (MoE/MoRA)?
2. To what extent do pre-service EFL teachers feel that their current teacher education programmes prepare them for classroom assessment?
3. What recommendations do pre-service EFL teachers have for their current tutors regarding strengthening LAL on their current teacher education programmes?

Methods

Design and Participants

A sequential explanatory design was employed, gathering data in two consecutive stages: the collection and analysis of quantitative data, followed by the collection and analysis of qualitative data. All participants were pre-service EFL teachers in their final year of undergraduate teacher education programmes across Indonesia. In the quantitative stage, 320 participants (108 males and 212 females) were selected using convenience sampling, allowing for the examination of the complex concept under investigation. In the qualitative stage, 37 participants (17 males and 20 females) were recruited for semi-structured interviews through snowball sampling (in which existing participants referred new participants), providing an in-depth exploration of the quantitative findings. The 320 participants represented both private ($n = 104$) and state ($n = 216$) universities, with affiliations to the Ministry of Education (MoE) ($n = 216$) and the Ministry of Religious Affairs (MoRA) ($n = 104$).

Instruments and Procedure

The data were collected using two instruments: a questionnaire and a semi-structured interview. The pilot phase did not raise any issues; therefore, the same instruments were used in the study.

Quantitative Stage

This stage involved developing a questionnaire measuring pre-service EFL teachers' LAL and their levels of satisfaction with how their undergraduate teacher education programmes prepared them for their future jobs in terms of demonstrating assessment competence. This measure was used to answer the first and second research questions. It was based on previous LAL surveys, including those conducted by DeLuca and Klinger (2010), Fulcher (2012), and Vogt and Tsagari (2014). Specifically, the part on teachers' LAL levels consisted of two subscales: pre-service teachers' theoretical knowledge of language assessment (34 items) and their ability to implement assessment in language classrooms (47 items). The latter consisted of three components, namely, teachers' ability to plan and design assessments (20 items), administer and score assessments (17 items), and utilise assessment results (10 items). The questionnaire used six-point Likert scales ranging from 1 being "not at all prepared" to 6 being "highly prepared". Both descriptive and inferential statistics were employed. The latter took the form of an independent samples t-test to examine possible differences in LAL across university status and affiliations.

The reliability of the questionnaire, measured using Cronbach's Alpha, was 0.9, indicating a very high level of internal consistency. To enhance content validity, the instrument was reviewed by two experts in the field of language assessment, one local and one international. Pre-service teachers' levels of LAL were determined using the following scale: 2.66 or lower = low, 2.67 to 4.32 = medium, and 4.33 to 6.00 = high.

Qualitative Stage

A semi-structured interview was employed in the qualitative stage. This helped further explore pre-service teachers' satisfaction with how their undergraduate teacher education programmes prepared them for demonstrating assessment competence in their future jobs (the second research question). Additionally, it was utilised to gather data on pre-service teachers' recommendations for strengthening LAL on their current teacher education programmes (the

third research question). Specifically, the purpose of the interview stage was threefold: to clarify the information provided through the questionnaire, to obtain a certain amount of consistent information from all the interviewees, and to allow for elaboration applicable to all interviewees through follow-up questions. The open-ended questions, among others, included: *How confident do you feel about assessing your future students? What specific aspects of assessment were NOT well covered on your undergraduate teacher education programme? What suggestions for improving the assessment component of your undergraduate teacher education programme do you have for your lecturers/universities?* The participants were interviewed separately to avoid any external influence. Each interview lasted 30 to 45 minutes, depending on how engaged the participants were and how much information they were willing to share. The participants could use either English or Bahasa Indonesia at the interview stage – whichever made them more comfortable. The transcripts in Bahasa Indonesia were then translated into English.

Thematic coding was then employed to “identify, analyse, and report [themes] within data” (Braun & Clarke, 2006, p. 79). In other words, the qualitative analysis focused on the potential of the participants’ narratives to explain and facilitate a deeper understanding of the studied phenomena (Swain et al., 2011). To achieve this, it followed Braun and Clarke’s (2013, pp. 202–203) six stages of thematic analysis: familiarisation with the data, generating initial codes, developing themes from codes, reviewing themes, defining and naming themes, and writing up the results.

Ethical Considerations

The study was guided by ethical principles established by the British Educational Research Association (2018). Confidentiality and anonymity were maintained throughout. The participants were asked to sign consent forms before the study, which clarified their right to withdraw from the project. Ethical approval for this study was obtained from the Ethics Committee at a local university in Indonesia. Upon completion of the qualitative stage, member checking was employed to explore the credibility of the results (Guba & Lincoln, 1989; Hays & Singh, 2012). The participants were given the opportunity to review their interview transcriptions for accuracy and alignment with their experiences.

Results

This section analyses the data in three separate subsections. The first subsection presents quantitative data to answer the first research question. The second section uses both quantitative and qualitative data to answer the second research question. The last section focuses purely on qualitative data and seeks to answer the third research question.

Levels of LAL among Indonesian Pre-Service EFL Teachers: Correlations with University Status (Private/State) and Institutional Affiliation (MoE/MoRA)

The purpose of this section is to answer the first research question: *What are the existing levels of LAL among Indonesian pre-service EFL teachers, and do these levels vary according to university status and university affiliation?* Consequently, descriptive statistics in mean scores and standard deviations were calculated in relation to the four components of LAL: theoretical knowledge, ability to plan and design assessments, ability to administer and score assessments, and ability to make use of assessment results (Tables 1 to 4).

Table 1
Pre-Service Teachers' Theoretical Knowledge of Language Assessment

No	Items	M	SD
	My knowledge of:		
1	the difference between assessment, testing, and evaluation	4.05	1.06
2	types of language assessment tasks	4.47	1.15
3	formative assessment	4.23	1.13
4	summative assessment	4.23	1.14
5	proficiency tests	4.02	1.03
6	achievement tests	4.04	1.13
7	diagnostic tests	4.02	1.11
8	placement tests	3.88	1.09
9	portfolio-based assessment	3.95	1.01
10	log-based assessment	3.36	1.07
11	self-assessment	4.36	1.15
12	peer-assessment	4.05	1.18
13	criterion-referenced assessment	3.53	1.22
14	norm-referenced assessment	3.35	1.18
15	the concept of reliability	3.71	1.09
16	the concept of validity	3.94	1.11
17	the concept of practicality	3.71	1.09
18	the concept of authenticity	4.02	1.14
19	the concept of washback	3.69	1.09
20	assessing listening skills	4.28	1.11
21	assessing reading skills	4.37	1.08
22	assessing speaking skills	4.31	1.07
23	assessing writing skills	4.39	1.03
24	assessing grammar	4.10	1.07
25	assessing pronunciation	4.28	1.05
26	assessing vocabulary	4.37	1.11
27	assessing intercultural competence	3.73	1.00
28	assessing non-verbal communication	3.85	1.19
29	Indonesian standards and policies of assessment	3.70	1.10
30	the Common European Framework of Reference	3.22	1.17
31	using technology in language assessment	4.35	1.11
32	assessing young learners	4.16	1.07
33	assessing adolescent learners	3.80	1.05
34	assessing adult learners	3.87	1.04
	Average	3.98	0.82

Table 2
Pre-Service Teachers' Ability to Plan and Design Assessments

No	Items	M	SD
	I can:		
1	identify the purposes of assessments	4.15	1.08
2	align assessments to national standards	3.77	1.07
3	align assessments to the Common European Framework of Reference	3.12	1.08
4	develop assessment plans/specifications for summative assessments	3.79	1.05
5	design assessments reflecting course content	3.80	1.06
6	design assessments resembling real-life situations	3.97	1.03
7	design criterion-referenced assessments	3.50	1.13
8	design different types of assessment tasks aligned with learning objectives	3.91	1.11
9	provide clear instructions for assessment tasks	4.20	1.13
10	determine the number of items or tasks within a test	4.02	1.12
11	develop scoring rubrics for assessing language skills	3.90	1.05
12	develop scoring keys for assessing language areas	3.87	1.06

13	allocate marks to individual tasks in a test	4.01	1.08
14	allocate timing to individual tasks in a test	3.95	1.08
15	sequence tasks within a test	3.78	1.05
16	design reliable tests	3.76	1.11
17	design valid tests	3.77	1.10
18	conduct item analysis in tests	3.67	1.11
19	decide on the layout of tests	3.68	1.12
20	design technology-based assessments	4.26	1.21
Average		3.84	0.88

Table 3
Pre-Service Teachers' Ability to Administer and Score Assessments

No	Items	M	SD
I can:			
1	use formal observation to assess students' progress	3.88	1.05
2	encourage students to self-assess their work using a range of tools	3.79	1.06
3	encourage students to self-assess their work using a range of tools	4.02	1.04
4	promote portfolio-based assessments	3.71	1.13
5	ensure that assessments take place on previously set dates	3.81	1.06
6	ensure that assessments take place in conducive environments	3.98	1.09
7	ensure that assessments abide by school regulations	3.89	1.11
8	invigilate assessments professionally	4.00	1.11
9	ensure that students' personal characteristics do not affect assessment scores	4.18	1.29
10	assess students fairly	4.48	1.18
11	assess students consistently	4.38	1.17
12	ensure the confidentiality of students' scores	4.28	1.16
13	ensure the anonymity of students' scores	4.21	1.17
14	use scoring rubrics to assess language skills	4.13	1.19
15	use statistics to analyse assessment scores	3.70	1.19
16	assess students with disabilities or learning difficulties	3.65	1.17
17	provide effective feedback	3.94	1.14
Average		4.00	0.92

Table 4
Pre-Service Teachers' Ability to Make Use of Assessment Results

No	Items	M	SD
I can use assessment results to:			
1	motivate students to learn	4.38	1.19
2	award prizes to students	4.34	1.16
3	diagnose students' strengths and weaknesses	4.25	1.22
4	compare students' progress with national standards	4.01	1.19
5	reflect on the teaching-learning process to improve it	4.16	1.21
6	evaluate the effectiveness of my language course	4.16	1.18
7	provide relevant support to students	4.16	1.19
8	place students in appropriate proficiency-level classes	4.18	1.10
9	inform the school of students' progress and achievements	4.20	1.18
10	inform parents of students' progress and achievements	4.31	1.08
Average		4.21	1.05

The overall mean scores suggested that the participants demonstrated moderate levels of LAL ($M = 4.01$; $SD = 0.69$ on a 6-point scale). Regarding the four components of LAL, the participants rated themselves higher on using assessment results ($M = 4.21$; $SD = 1.05$). Regarding the other three components, they perceived themselves as being reasonably high in the ability to administer and score assessments, in their theoretical knowledge of LAL, and in their ability to plan and

design assessments, with mean scores of 4.00 (SD = 0.92), 3.98 (SD = 0.82), and 3.84 (SD = 0.88), respectively.

Additionally, a series of independent samples t-tests was conducted to examine whether there were any statistically significant differences in the overall LAL score and the scores for each component of LAL among the teachers across university status and affiliations. Significant differences in overall LAL were identified between teachers enrolled in state and private universities ($t(318) = -3.983, p = .000$, effect size = 0.05), with results favouring state university attendees. The teachers from state universities also differed significantly from those from private universities in relation to their theoretical knowledge of language assessment ($t(318) = -4.287, p = .00$, effect size = 0.05), their ability to plan assessment in the language classroom ($t(318) = -3.655, p = .00$, effect size = 0.04), and their ability to administer and score assessments ($t(318) = -4.275, p = .00$, effect size = 0.05). The magnitude of the effect of university status was small for all components of LAL. However, no significant difference was observed in teachers' ability to use assessment results ($t(318) = .006, p = .99$, effect size = 0.00).

Regarding university affiliation, the independent samples t-test indicated that teachers from universities affiliated with the MoE exhibited higher levels of LAL in relation to theoretical knowledge of language assessment ($t(318) = 6.524, p = .00$, effect size = 0.12), ability to plan assessment in the language classroom ($t(318) = 6.384, p = .00$, effect size = 0.11), administering and scoring assessments ($t(318) = 5.763, p = .02$, effect size = 0.09), and making use of assessment results ($t(318) = 2.004, p = .04$, effect size = 0.01) than their counterparts from universities under the MoRA. Regarding the overall score, teachers in universities affiliated with the MoE also displayed a higher rate of LAL than their MoRA university counterparts ($t(318) = 6.762, p = .00$, effect size = 0.12). Unlike university status, the magnitude of the effect of university affiliation on the overall score and almost all components of LAL was moderate. The only exception was the final component of LAL – using assessment results.

Indonesian Pre-Service EFL Teachers' Perceptions of the Effectiveness of their Undergraduate Programme in Preparing them for Classroom Assessment

To answer the second research question (*To what extent do pre-service teachers feel their current teacher education programmes prepare them for classroom assessment?*), quantitative and qualitative data were gathered through questions in the questionnaire and interview stages.

Overall, most participants (81% in the quantitative stage and 91% in the qualitative stage) were satisfied with their teacher education programmes. The quantitative stage revealed that they liked the fact that assessment aspects were integrated into the curriculum, either as separate courses on assessment or as parts of other courses of a pedagogical nature, and were clearly explained in general. On closer inspection, 92% of the respondents from universities under the MoE (70% from state universities and 22% from private universities) stated that the sessions on assessment were both theoretical and practical. This alignment with expectations contributed to an enhancement in the overall assessment literacy of future teachers. However, slightly more than 50% of the respondents in both state and private contexts asked for more focus on practical classroom applications. This resonates with the qualitative stage, where two interviewees expressed this view as follows:

I enjoyed the assessment course because it was both theoretical and practical. I learnt a lot, but I think my theoretical understanding of assessment is better than my practical ability to implement it. I think my lecturer should have devoted a bit more time to practical tasks so I would feel more confident in my teaching job. (Interviewee 17)

Although I liked my sessions on assessment...more attention should have been given to practical tasks, such as using portfolios or designing rubrics, so I could...deal with them in my teaching. There was too much theory...more focus on the complex Indonesian education system would have been helpful. (Interviewee 27)

Regarding the respondents from MoRA universities, 63% (30% from state universities and 33% from private universities) reported that the sessions on assessment were mainly theoretical with hardly any practical activities, which begs for more pragmatic approaches to integrating LAL into teacher education programmes. Similarly, three interviewees stated the following:

My sessions on assessment were full of theory. My lecturer did not focus on classroom practice. Although I understand many concepts, I do not think I am well prepared for assessing students in the classroom. More practical training needs to be provided, and pre-service teachers' awareness of assessment competence and its importance needs to be raised. (Interviewee 18)

Our sessions on assessment focused mainly on language testing and its reliability and validity. Lectures were delivered, and some reading was assigned, but these sessions were not prepared with future teachers in mind; there was no practice. I also thought assessment was broader than testing, but we never went beyond testing... (Interviewee 2)

My course was somewhat disappointing because it focused only on assessment theory. Our lecturer delivered...sessions on assessment, interesting ones...but did not give us any practical tasks. Some of these sessions were general in scope and therefore not very useful for language teaching. Language teaching and assessment differ significantly from other school subjects. (Interviewee 33)

Indonesian Pre-Service EFL Teachers' Recommendations for Strengthening LAL on Current Teacher Education Programmes

This section aims to answer the final research question: *What recommendations do pre-service teachers have for their current tutors regarding strengthening LAL on their current teacher education programmes?* Unlike the previous question, all participants approached the third question with a high degree of reflection, providing useful recommendations in the form of textual data. In the former, participants' responses tended to be influenced by their overall impressions of their programmes, not solely the LAL aspect, resulting in a relatively high outcome. Due to space limitations, only the three most prevalent recommendations are delineated below. These recommendations were consistently identified by interviewees from both state and private universities, as well as those governed by the MoE and MoRA.

Firstly, the interviewees thought there should be greater emphasis on the practical aspects of assessment. For instance, they wished to be involved in designing tests and rubrics to assess oral and written assignments and be trained to provide effective feedback. The latter was mentioned on several occasions; apparently, teacher education programmes prioritised summative assessment, thus marginalising assessment for learning. Two interviewees expressed their opinions as follows:

It is important that future teachers can design quality tests. Therefore, the course on assessment should not only clarify the step-by-step procedure, but also give pre-service teachers the opportunity to design tests in collaboration with their peers to put theory into practice. However, space for less traditional formats of assessment must also be created, as well as for ongoing classroom assessment. (Interviewee 6)

The current course should train teachers to give constructive feedback to students. Teachers must ensure it is balanced and focuses on positive and negative aspects. The latter are tricky to communicate. Therefore, future teachers must practise not to demotivate...their students. (Interviewee 11)

Secondly, the interviewees thought more attention should be paid to alternative assessment types. They emphasised that there was a tendency to focus on tests, but tests were only one of the ways of assessing students' progress. They believed that Indonesian teachers' attention should be drawn to project-based assignments, problem-based assignments, case-based scenarios, and portfolios. These types of assessments would be useful in assessing twenty-first-century skills widely promoted by current policies in Indonesia. For instance, one interviewee stated that:

Current courses on assessment should promote alternative assessments, as there are still too many tests in schools. For example, projects should be promoted...to familiarise teachers with holistic assessment where students' personal, social, and emotional development can be considered, including students' diverse needs and the local context the students come from. (Interviewee 22)

Thirdly, the interviewees believed that the courses on assessment should have provided more information on using technology and Artificial Intelligence (AI) in designing assessments and grading student work or identifying AI-generated work submitted by students. They thought that the technology aspect was generally missing on their programmes and, therefore, should be integrated into future courses on assessment. One interviewee explained that:

My course did not include technology at all. More training needs to be provided on how technology could be used in assessment in general or how AI could be used in designing assessments more specifically to make teachers' lives easier... Indonesian teachers' workloads are heavy, and their professional lives are pretty bureaucratic. (Interviewee 28)

Discussion

This section discusses the findings and is organised into three sections corresponding to the research questions. To maintain consistency, this structure aligns with the analysis presented in Section 4.

Levels of LAL and their Correlations with University Status and Affiliation

The quantitative data revealed that Indonesian pre-service EFL teachers possess a moderate level of LAL, with an overall mean score of 4.01 on a six-point scale. The highest competence was observed in using assessment results, suggesting that these teachers are confident in utilising assessment data to inform instruction, motivate students, and communicate progress. This competence is critical for effective teaching and learning, as it ensures that assessments serve not only as evaluative tools, but also as mechanisms for continuous improvement and student engagement in the teaching-learning process. This result contrasts with numerous studies of in-service teachers who have reported relatively low levels of LAL (e.g., Kiomrs et al., 2011; Mede & Atay, 2017; Tsagari & Vogt, 2014). For instance, Fitriyah et al. (2022) found that teachers in Indonesia only exhibited a functional or basic level of LAL. Conversely, Zulaiha et al.'s (2020) study reported a moderately high level of LAL among a group of junior secondary Indonesian EFL teachers. Given these findings, one might expect pre-service teachers to exhibit a more functional level of LAL. However, the results here do not necessarily indicate the development of more procedural or multidimensional LAL, as outlined in Pill and Harding's (2013) continuum.

The data also indicated variability in LAL based on university status and affiliation. Teachers from state universities demonstrated significantly higher overall LAL than their counterparts from private universities. This trend was consistent across the first three components of LAL (theoretical knowledge, planning and designing assessments, and administering and scoring assessments) but not in the ability to use assessment results – the final component of LAL in the administered questionnaire. The effect sizes for these differences were small, indicating that whilst

the differences were statistically significant, they might not have been substantial in practical terms.

Similarly, university affiliation (MoE vs. MoRA) moderately affected LAL. Teachers from universities under the MoE reported higher levels of LAL across all four components than those affiliated with the MoRA. This, however, should not be surprising as state universities enrol higher-quality students due to a more competitive recruitment process. Additionally, this disparity is likely to be attributable to the varying emphasis and resources allocated to teacher education programmes by the different ministries. The moderate effect size suggests a more pronounced practical significance compared with university status, highlighting a potential area for policy intervention to ensure more uniform teacher education standards, course content, and resource allocations across different affiliations.

Effectiveness of Undergraduate Programmes in Preparing Pre-Service EFL Teachers for Classroom Assessment

The quantitative and qualitative data provided detailed insights into pre-service teachers' perceptions of the effectiveness of their undergraduate programmes in preparing them for classroom assessment. A significant majority (81% in the quantitative stage and 91% in the qualitative stage) expressed satisfaction with their programmes, appreciating their focus on the overall enhancement of future teachers' assessment literacy.

However, the depth of satisfaction varied. Teachers from state universities and those under the MoE reported a more balanced approach, combining theoretical and practical aspects of assessment. Nevertheless, more than half of these respondents expressed a need for an increased practical focus and for more hands-on training where teachers could be involved in designing test questions, performance assessment tasks, and marking schemes, linking assessments to learning outcomes, and giving and receiving feedback. By contrast, teachers from private universities and those under the MoRA noted a predominantly theoretical approach with minimal practical application. Their feedback emphasised the inadequacy of practical training, which frequently left them feeling underprepared for real-world classroom assessment tasks. They believed that this gap between theoretical knowledge and practical application was critical, as it directly impacts teachers' confidence and competence in preparing, administering, and grading assessments effectively.

In the first instance, this finding may support Fulcher's (2020) apprenticeship model of cycles of test development in local contexts. Concurrently, studies have demonstrated that both in-service and pre-service teachers express a need for more practical forms of training in assessment and testing rather than theoretical approaches (e.g., Giraldo & Murcia, 2018; Yan et al., 2018). In the Indonesian context, university teachers in Puspawati's study (2019) expressed greater confidence in their knowledge of principles but significantly less confidence in their skills. This finding may also echo Berry et al.'s (2017) study, which highlighted teachers' belief in the primacy of experience and practice in developing LAL.

Recommendations for Strengthening LAL on Teacher Education Programmes

The participants also provided several recommendations for enhancing the LAL component of their teacher education programmes. The most prominent suggestion was to increase the practical emphasis within courses on assessment. The interviewees expressed a desire for more opportunities to design tests, develop rubrics, and practise providing effective feedback. These activities are crucial for translating theoretical knowledge into practical skills, thereby enhancing

pre-service teachers' preparedness for classroom assessment and addressing the associated challenges and realities within the Indonesian context.

Additionally, there was a call for greater attention to alternative assessment types. Introducing project-based, problem-based, and portfolio assessments could provide a more comprehensive evaluation of student progress by considering their diverse needs, learning preferences, multiple intelligences, abilities, and interests. The current focus on traditional tests does not fully equip teachers to assess twenty-first-century skills, which are crucial for holistic student development. This finding resonates with Vogt and Tsagari's (2014) interview-based study of secondary EFL teachers in Europe, who perceived that traditional test approaches predominated over the more informal formative assessment activities that teachers needed. Interestingly, there was no specific call from Indonesian pre-service teachers for a greater focus on pedagogical methods and language principles, as highlighted in previous research on needs (e.g., Fulcher, 2020; Giraldo, 2021). It appears that a realisation of these needs may only come with the experience that the Indonesian participants lacked.

Finally, the integration of technology and AI into assessment practices was highlighted as a critical need, aligning with Owan et al.'s (2023) deliberations on the potential of artificial intelligence tools in educational assessment. Given the expanding role of technology and AI in language education and educational measurement in general, it is imperative that pre-service teachers be adept at using digital tools for assessment purposes. This proficiency encompasses designing and grading assessments, identifying AI-generated work, utilising digital portfolios, employing computer-based testing platforms, and leveraging AI to create innovative assessment formats (Owen et al., 2023). However, while the advantages of technology- and AI-supported assessment are evident, teachers must remain cognisant of the new challenges these advancements bring. The lack of adequate training in current teacher education programmes highlights a significant gap that must be addressed to effectively equip teachers with the skills necessary to assess students' knowledge, skills, and competences in twenty-first-century schools (Fulcher, 2020).

Conclusion

The purpose of the current study was threefold: (1) to measure the level of LAL among Indonesian pre-service EFL teachers, ascertaining whether their levels of LAL varied according to university status and university affiliation, (2) to explore whether current undergraduate teacher education programmes prepared pre-service EFL teachers for classroom assessment, and (3) to identify recommendations pre-service teachers had for their tutors with regard to strengthening teacher LAL on their current teacher education programmes.

The findings offer valuable insights into the current state of LAL among Indonesian pre-service EFL teachers. As the analyses show, the overall levels of LAL among these teachers are moderate, with notable variations contingent on university status and affiliation. The qualitative data from the participants underscore the need for a more practical focus in assessment training, greater emphasis on alternative assessment methods, and the integration of technology and AI. Addressing these recommendations will help bridge the gap between theoretical knowledge and practical application, ultimately enhancing the effectiveness of teacher education programmes in preparing competent and confident EFL teachers.

The study has several limitations that should be addressed in future research. Firstly, the use of convenience and snowball sampling may have introduced selection bias, affecting the generalisability of the findings. The reliance on self-reported data could have introduced response bias, with participants possibly overestimating their LAL or programme satisfaction levels.

Additionally, the sample consisted solely of pre-service EFL teachers, limiting the applicability of the results to other educational levels, such as postgraduate teacher education programmes. Although the instruments were piloted and demonstrated high reliability, the validity of the interview responses could have been influenced by varying levels of English proficiency among participants. These limitations suggest that, while the study provides valuable insights into pre-service EFL teachers' LAL in Indonesia, caution should be exercised in generalising the findings beyond this specific context.

The findings have significant implications for preparing pre-service EFL teachers in Indonesia and Southeast Asia. Teacher educators should incorporate more practical, hands-on experiences in assessment practices to bridge the gap between theory and classroom practice. Expanding the focus to include alternative assessments and the use of technology and AI would better prepare teachers for diverse classroom demands. These changes could lead to more confident and competent teachers who use formative and summative assessment effectively. Policymakers must ensure that teacher education programmes offer comprehensive assessment training. Additionally, variability based on university status underscores the need for standardised training across institutions to reduce gaps between state and private, and MoE and MoRA institutions.

References

- Bachman, L., & Damböck, B. (2018). *Language assessment for classroom teachers*. Oxford University Press.
- Berry, V., Sheehan, S., & Munro, S. (2017). Exploring teachers' language assessment literacy: A social constructivist approach to understanding effective practices. In ALTE (Ed.), *Learning and assessment: Making the connections* (pp. 201–207). *Proceedings of the ALTE 6th International Conference*. Association of Language Testers in Europe.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7–74. <http://dx.doi.org/10.1080/0969595980050102>
- British Educational Research Association. (2018). *Ethical guidelines for educational research* (4th ed.). British Educational Research Association. <https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018-online>
- Cirocki, A., & Hallet, W. (2024). *Professionalising English language teaching: Concepts and reflections for action in teacher education*. Cambridge University Press.
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(3). <https://link.springer.com/content/pdf/10.1186/s40468-020-00101-6.pdf>
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://journals.sagepub.com/doi/10.1177/0265532208090156>

- Davison, C., & Leung, C. (2012). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415. <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1545-7249.2009.tb00242.x>
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419–438. <https://www.tandfonline.com/doi/abs/10.1080/0969594X.2010.516643>
- DeLuca, C., Coombs, A., MacGregor, S., & Rasooli, A. (2019). Toward a differential and situated view of assessment literacy: Studying teachers' responses to classroom assessment scenarios. *Frontiers in Education*, 4. <https://doi.org/10.3389/feduc.2019.00094>
- Directorate of Teacher Profession Education. (2024). **2024 pre-service PPG policy direction**. <https://ppg.kemdikbud.go.id/prajab/detail-program-kurikulum>
- Fitriyah, I., Bastomi, Y., Khotimah, K., & Gozali, I. (2022). Implementation of assessment for learning in online EFL writing class: A case of novice undergraduate teachers. *LEARN Journal: Language Education and Acquisition Research Network*, 15(2), 129–159. <https://files.eric.ed.gov/fulltext/EJ1358706.pdf>
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Fulcher, G. (2020). Operationalizing assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 8–28). Cambridge Scholars Publishing.
- Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *HOW*, 26(1), 35–61. <https://www.howjournalcolombia.org/index.php/how/article/view/481/499>
- Giraldo, F. (2021). Language assessment literacy and teachers' professional development: A review of the literature. *Profile: Issues in Teachers' Professional Development*, 23(2), 265–279. <https://doi.org/10.15446/profile.v23n2.90533>
- Giraldo, F. (2023). *Fostering pre-service teachers' language assessment literacy*. Editorial Ucaldas.
- Giraldo, F., & Murcia, D. (2018). Language assessment literacy for pre-service teachers: Course expectations from different stakeholders. *GIST – Education and Learning Research Journal*, 16, 56–77. <https://doi.org/10.26817/16925777.425>
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Sage.
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 413–428). De Gruyter.
- Hay, P., & Penney, D. (2013). *Assessment in physical education: A sociocultural perspective*. Taylor & Francis.
- Hays, D. G., & Singh, A. A. (2012). *Qualitative inquiry in clinical and educational settings*. Guilford.

- Hildén, R., & Fröjdendahl, B. (2018). The dawn of assessment literacy – Exploring the conceptions of Finnish student teachers in foreign languages. *Apples: Journal of Applied Language Studies*, 12, 1–24. <https://doi.org/10.17011/apples/urn.201802201542>
- 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>
- Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301–307.
<https://journals.sagepub.com/doi/10.1177/0265532213480126>
- Inbar-Lourie, O. (2016). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 1–14). Springer.
- 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. <https://doi.org/10.1080/09695940903075958>
- Kiomrs, R., Abdolmehdi, R., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching*, 4(1), 156–161.
<https://ccsenet.org/journal/index.php/elt/article/view/9677>
- Kremmel, B., & Harding, L. (2019). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100–120.
<https://doi.org/10.1080/15434303.2019.1674855>
- Levi, T., & Inbar-Lourie, O. (2019). 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>
- McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction*. Pearson.
- Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi*, 168(1), 43–60. <https://www.cceol.com/search/article-detail?id=548906>
- Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montréal, Quebec, Canada.
- Ministry of Education, Culture, Research and Technology. (2022). *Regulation of Ministry of Education, Culture, Research and Technology, No 21 2022, on Assessment standard for early childhood education, primary and secondary school education*. Ministry of Education, Culture, Research and Technology.

- Nyudak, I. B. G., Dewi, N. L. P. E. S., & A.A. Gede Yudha Paramartha, A. A. G. Y. (2022). Assessment literacy of EFL teachers in Badung, Bali, Indonesia: Conception and practices. *ELT Worldwide: Journal of English Language Teaching*, 9(1), 178–187.
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>
- Papageorgiou, S., & Bailey, K. M. (Eds.). (2019). *Global perspectives on language assessment: Research, theory, and practice*. New York: Routledge.
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128–138. <https://doi.org/10.1016/j.tate.2019.05.003>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381–402. <https://doi.org/10.1177/0265532213480337>
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Puspawati, I. (2019). Understanding teachers' knowledge, skills, and principles of language assessment: A survey on teachers' language assessment literacy. <https://www.academia.edu/100883188>
- Putra, I., Paramartha, A., & Mahayanti, N. (2024). Online assessment literacy level of EFL teachers in Denpasar. *Measurement in Educational Research*, 4(1), 49–56. <http://dx.doi.org/10.33292/meter.v4i1.321>
- Rezagah, K. (2022). Iranian EFL teachers' assessment literacy knowledge: The impact of the educational system on teachers' classroom assessment practices. *International Journal of Language Testing*, 12(2), 76–94. <https://doi.org/10.22034/IJLT.2022.157126>
- Sevimel-Sahin, A. (2020). A survey on the intercultural communicative competence of ELT undergraduate students. *Eurasian Journal of Applied Linguistics*, 6(2), 141–153. <https://doi.org/10.32601/ejal.775793>
- Shim, K. N. (2009). *An investigation into teachers' perceptions of classroom based assessment of English as a foreign language in Korean primary education*. Doctoral Thesis. University of Exeter.
- Stiggins, R. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement Issues and Practice*, 18, 23–27. <https://doi.org/10.1111/j.1745-3992.1999.tb00004.x>
- Suherman, A. (2022). Exploring language assessment literacy of EFL teachers in the context of Indonesian higher education. *English Review: Journal of English Education*, 10(3), 783–792. <https://doi.org/10.25134/erjee.v10i3.6525>
- Swain, M., Kinnear, P., & Steinman, L. (2011). *Sociocultural theory in second language education*. Multilingual Matters.

- Tajeddin, Z., Alemi, M., & Yasaci, H. (2018). Classroom Assessment Literacy for Speaking: Exploring Novice and Experienced English Language Teachers' Knowledge and Practice. *Iranian Journal of Language Teaching Research*, 6 (Issue 3 (Special Issue)), 57-77. <https://doi.org/10.30466/ijltr.2018.120601>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. <https://doi.org/10.1177/0265532213480338>
- Tsagari, D. (2020). Language assessment literacy: Concepts, challenges and prospects. In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning*. Routledge.
- Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63. <https://doi.org/10.58379/UHIX9883>
- UCLES. (2018). *Cambridge English teaching. Framework competency statements*. UCLES.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-392. <https://doi.org/10.1080/15434303.2014.960046>
- Weng, F., & Shen, B. (2022). Language assessment literacy of teachers. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.864582>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Yan, X., Zhang, C., & Fan, J. J. (2018). "Assessment knowledge is important, but ...": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158-168. <https://doi.org/10.1016/j.system.2018.03.003>
- Yeo, M. A. (2021). "Experiencing theory first-hand was delightful and informative": Bridging the theory-practice gap in online language assessment training. *Iranian Journal of Language Teaching Research*, 9(3), 93-116. <https://doi.org/10.30466/ijltr.2021.121078>
- Zulaiha, S., Mulyono, H., & Ambarsari, L. (2020). An investigation into EFL teachers' assessment literacy: Indonesian teachers' perceptions and classroom practice. *European Journal of Contemporary Education*, 9(1), 189-201. <https://doi.org/10.13187/ejced.2020.1.189>

Andrzej Cirocki is a teacher educator and ELT materials developer. In addition to authoring numerous publications, participating in international research projects, and delivering CPD courses for TESOL teachers worldwide, he serves as the editor-in-chief of *The European Journal of Applied Linguistics and TEFL*. ORCID: <https://orcid.org/0000-0003-0442-5092>

Syafi'ul Anam teaches courses in academic writing, language assessment, materials development, and research methods. His primary research interests include self-regulated learning, self-efficacy beliefs, and assessment in EFL contexts. ORCID: <https://orcid.org/0000-0002-4687-856X>

Nur Arifah Draji is a teacher educator. Her research interests include technology in language learning, TPACK, and multimodal and informal digital learning of English. She regularly publishes in her area of expertise and contributes to English language teacher professional development in Indonesia. ORCID: <https://orcid.org/0000-0001-5828-5107>

Bill Soden is a teacher educator and teacher trainer. He has published work on English for Academic Purposes, postgraduate student academic writing, assessment and feedback in higher education, and, most recently, testing and assessment in the ELT classroom. ORCID: <https://orcid.org/0000-0002-7621-2320>