

Exploring Classroom-Based Speaking Assessment Literacy Among Novice English Teachers: Assessment Knowledge and Classroom Enactment

Sundari Purwaningsih^{1*}

¹Teacher Training and Education Faculty, Universitas Perjuangan Tasikmalaya, Indonesia
(sundaripurwaningsih@unper.ac.id)

Lystiana Nurhayat Hakim²

²Teacher Training and Education Faculty, Universitas Perjuangan Tasikmalaya, Indonesia
(lystiananurhayat@unper.ac.id)

Tri Agustini Solihati³

³Teacher Training and Education Faculty, Universitas Perjuangan Tasikmalaya, Indonesia
(triagustini@unper.ac.id)

**corresponding author*

Abstract: *This study explores classroom-based speaking assessment literacy among novice English teachers who had completed Indonesia's PPG Prajabatan program, focusing on assessment knowledge and its relationship to classroom enactment. A qualitative multiple-case study was conducted in two phases. In Phase 1, ten teachers completed scenario-based tasks concerning assessment purposes, content/criteria, and methods, and seven participated in interviews. In Phase 2, classroom observations, follow-up interviews, and instructional documents were used to examine the practices of three focal teachers. Data were analyzed using the Teacher Assessment Literacy in Practice (TALiP) framework. The findings show that eight participants articulated formative purposes and six articulated summative purposes. Pronunciation and fluency were the most frequently identified criteria, while presentation or monologic performance was the most frequently articulated assessment method. In the three observed classrooms, speaking assessment served both learning and grading purposes, indicating broad correspondence in assessment purposes. However, criteria documented in lesson plans were not explicitly communicated to students during the observed speaking tasks, indicating a planning–practice discrepancy. Oral performance tasks were used in practice, but teachers relied substantially on scripted and rehearsed performances, which they described as necessary for learners with basic English proficiency and as manageable within classroom constraints. This reliance limited opportunities to observe more spontaneous oral language use. Overall, the relationship between assessment knowledge and classroom enactment varied across assessment domains and was context-dependent.*

Keywords: *speaking assessment; classroom-based speaking assessment literacy; novice English teachers; teacher assessment literacy in practice*

Article History:

Received : 18 December 2025
Revised : 29 September 2026

Accepted : 29 September 2026
Available Online : 30 September 2026



Suggested Citation Format:

Purwaningsih, S., Hakim, L.N., Solihati, T.A. (2026). Exploring Classroom-Based Speaking Assessment Literacy Among Novice English Teachers: Assessment Knowledge and Classroom Enactment. *PANYONARA: Journal of English Education*, 8(2), 141–173.
<https://doi.org/10.19105/panyonara.v8i2.23250>

INTRODUCTION

Assessment is integral to language teaching because assessment decisions can influence classroom instruction and students' learning experiences (Fulcher, 2021; Tsagari & Vogt, 2017). Assessment has also been associated with student achievement and motivation (Alderson et al., 2017; Coombs et al., 2018; Vogt et al., 2020). Its significance is further reflected in the time teachers devote to it; foundational research indicates that classroom assessment activities consume considerable time, with estimates reaching up to 50% of teachers' instructional time (Plake et al., 1993; Stiggins, 1991).

Assessment data are also frequently used by educational stakeholders to inform policy decisions (Fitriyah et al., 2022). Therefore, English language teachers must possess a set of competencies that integrate the perspectives of multiple stakeholders and encompass the knowledge, skills, and principles required for assessing language skills, commonly referred to as language assessment literacy (LAL) (Giraldo, 2018). A sufficient level of LAL enables teachers to design and implement assessment procedures that are valid and reliable (Giraldo, 2018; Koh et al., 2018), whereas limited LAL may compromise the reliability and validity of classroom assessments, thereby negatively influencing student learning outcomes (Xu & Brown, 2016).

Existing research does not point to a uniform level of language assessment literacy among teachers. Studies from different European and Asian settings have documented shortcomings in assessment knowledge and practice, but these vary according to the populations and contexts investigated (Lam, 2015; Lan & Fan, 2019; Qotboddin et al., 2020; Sultana, 2019; Tsagari & Vogt, 2017). The Indonesian evidence is particularly varied. Findings range from relatively adequate assessment literacy to moderate or low levels (Anam & Putri, 2021; Luthfiyyah et al., 2020a; Prastikawati et al., 2024; Suherman, 2022; Zulaiha et al., 2020). Such variation makes it difficult to characterize Indonesian English teachers' assessment literacy in terms of a single level. It instead points to the importance of examining what teachers know about assessment and how that knowledge operates within particular teaching contexts.

Classroom-based assessment (CBA) is a critical aspect of language teaching and learning (Derakhshan et al., 2020; Estaji & Ghiasvand, 2021). The Indonesian assessment context has undergone major changes in recent years. In 2021, Ujian Nasional (UN), the assessment system for individual student achievement, was replaced by Asesmen Nasional

(AN) to assess and map the quality of the education system rather than individual student achievement (Pusat Asesmen Pendidikan, 2021). At the classroom level, national assessment criteria focus on assessment as a means of gathering and processing information about students' learning needs and achievement (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, 2022). This policy context means that the ability of teachers to design and implement classroom assessment is important. Unlike large-scale summative assessments performed under highly controlled conditions, CBA is integrated into the instructional process and serves as an ongoing mechanism to support and guide student learning (Davison & Leung, 2009). This shift requires English language teachers to develop classroom-based assessment literacy (CBAL) so that they can design and implement classroom assessments that are not only measures of academic achievement but also foster deep learning (Shepard, 2019).

Research on CBAL has addressed both what teachers need to know about assessment and how that knowledge is used in classroom practice. Studies have examined teachers' assessment knowledge, the design and implementation of classroom assessment, and areas in which further professional development is needed (Gan et al., 2019; Koh et al., 2018; Lan & Fan, 2019). Collectively, this work also shows that assessment literacy cannot be reduced to technical competence in constructing and administering assessments. Teachers' assessment decisions are influenced by the instructional contexts in which they work, their beliefs about assessment, and the social and affective dimensions of classroom practice. This broader view is particularly evident in Xu and Brown's (2016) conceptualization of assessment literacy as situated practice rather than simply an individual body of assessment knowledge and skills.

Indonesian research has examined CBAL from several perspectives. Studies of practicing teachers have identified continuing professional-development needs in classroom assessment (Fitriyah et al., 2022; Zulaiha & Mulyono, 2020), while research involving both pre-service and in-service teachers has investigated their assessment knowledge, beliefs, and perceptions (Anam & Putri, 2021; Aria et al., 2021; Damayanti et al., 2024; Sumardi & Guci, 2023; Zulaiha et al., 2020). Preparation for assessment also appears to be a concern before teachers enter the profession. Hakim et al. (2024) reported that pre-service English teachers felt insufficiently prepared by their university coursework to design effective assessments. Together, these studies point to continuing concerns about how teachers develop assessment literacy, but they tell us less about how such knowledge is enacted in actual classroom assessment.

Assessing speaking performance presents unique challenges because it involves live interaction, requiring assessors to manage linguistic, paralinguistic, and pragmatic factors simultaneously under real-time classroom conditions (Fulcher, 2015). While research has begun examining domain-specific assessment literacy—such as Tajeddin et al.'s (2018) investigation of novice and experienced EFL teachers' speaking assessment knowledge and scenario practices—as well as secondary teachers' assessment literacy and conceptions in Indonesia (Damayanti et al., 2024; Luthfiyyah et al., 2020b), important empirical gaps remain. First, much of the prior research in Indonesia has relied on self-reported questionnaires or interviews, capturing self-perceived practices rather than live classroom enactment (Aria et al., 2021; Fitriyah et al., 2022; Luthfiyyah et al., 2020b). Where classroom observations have been employed, studies have examined general language assessment literacy among secondary EFL teachers rather than the skill-specific assessment demands of speaking (Damayanti et al., 2024). Second, little is known about how novice teachers who have completed Indonesia's national *PPG Prajabatan* or Pre-service Teacher Professional Education program operationalize their speaking assessment knowledge in practice. By combining scenario-based knowledge tasks with direct classroom observations of novice teachers who had completed *PPG Prajabatan*, this study examines how assessment knowledge is enacted in actual speaking assessment within secondary EFL classrooms.

To address these gaps, this study seeks to investigate the knowledge and practices of CBA for speaking skills among novice English teachers. Specifically, it seeks to answer the following questions:

1. What assessment knowledge do novice English teachers demonstrate in responding to scenarios concerning classroom-based speaking assessment in terms of purposes, content/criteria, and methods?
2. How does the assessment knowledge demonstrated by the observed novice English teachers relate to their classroom enactment of speaking assessment concerning purposes, content/criteria, and methods?

The findings are expected to advance understanding of classroom-based assessment literacy in the specific domain of speaking, and to provide evidence from actual classroom practice that can inform *PPG Prajabatan* curricula, teacher professional development, and efforts to support assessment practice among novice English teachers in Indonesia.

LITERATURE REVIEW

Language Assessment Literacy (LAL), Learning-Oriented Assessment (LOA), and Assessment Literacy in Practice

Assessment literacy (AL) has broadened considerably since Stiggins (1991) first associated it with teachers' ability to use assessment information in making pedagogical decisions. In language education, Davies (2008) conceptualized language assessment literacy in terms of three core components—knowledge, skills, and principles. Later work has expanded this technically oriented view by giving greater attention to the contexts in which assessment takes place (Baker, 2016; Crusan et al., 2016; Giraldo, 2018; Lan & Fan, 2019). Taylor's (2013) eight-dimensional model illustrates this development by incorporating dimensions that extend beyond technical assessment competence. Taken together, these developments show a shift from viewing assessment literacy primarily as a body of knowledge and skills toward understanding it as a competence shaped by the broader context of assessment practice.

Building on this socio-cultural shift, Inbar-Lourie (2008) further emphasized the importance of constructing a language assessment knowledge base that reflects the particular demands of language assessment rather than relying solely on generic testing knowledge. Similarly, Fulcher (2012) proposed an expanded definition of LAL that integrates technical knowledge, skills, and abilities with an understanding of guiding principles, ethics, and the broader historical, political, and socio-cultural frameworks that shape assessment impact. Recent empirical work has further demonstrated that LAL is non-linear and stakeholder-differentiated. Kremmel and Harding (2020) developed an empirical model comprising multiple dimensions of LAL and showed that different stakeholder groups have distinct assessment-related needs, with language teachers requiring practice-oriented competencies related to areas such as assessment in language pedagogy and washback. Yan and Fan (2021) likewise showed that LAL development is an ongoing, experience-mediated process shaped by local institutional contexts and professional experiences.

Within classroom settings, this socio-cultural evolution has aligned LAL closely with Learning-Oriented Assessment (LOA). LOA shifts attention toward the role of assessment in supporting and developing student learning rather than treating assessment solely as a means of measuring achievement. Carless (2015) articulates three interlocking principles of LOA: designing assessment tasks as authentic learning tasks, actively engaging students as co-assessors to build evaluative expertise, and providing timely feedback that functions as *feedforward* to guide future learning. Similarly, Turner & Purpura (2016) and Davison and Leung (2009) frame teacher-based assessment as an embedded process that integrates assessment with classroom teaching and learning. From this perspective, making assessment criteria accessible to learners and involving them in peer- or self-assessment can support their development of evaluative and self-regulatory capacities (Carless, 2015).

To capture how teachers navigate these complex classroom realities, this study adopts the Teacher Assessment Literacy in Practice (TALiP) framework (Xu & Brown, 2016). Unlike earlier technical models, TALiP conceptualizes assessment literacy as an enacted and situated practice shaped by social, affective, institutional, and professional contexts. Within the framework, the teacher knowledge base includes knowledge of assessment purposes, content and criteria, and methods, providing the analytical focus for the present study. The framework is particularly relevant to examining how teachers negotiate assessment decisions within the practical realities of their classrooms. Empirical studies adopting practice-oriented and contextualized perspectives further demonstrate the value of examining assessment knowledge alongside classroom decision-making and the conditions in which assessment is enacted. For example, Tajeddin et al. (2018) examined novice and experienced EFL teachers' speaking assessment knowledge and practices using scenarios and interviews framed by Xu and Brown's framework, while other studies have examined how teachers apply assessment knowledge to language-specific assessment tasks or develop classroom-based formative assessment literacy through professional learning (e.g. Levi & Inbar-Lourie, 2020; Li & Gu, 2023). These studies support the need to examine assessment literacy as both a knowledge base and a practice situated within particular instructional contexts.

Speaking Assessment in EFL Contexts

Despite the global emphasis on communicative competence, speaking assessment is often described in the literature as an underserved area of research compared to writing or general assessment (Tajeddin et al., 2018). Speaking is defined as the verbal use of language to communicate meaning through systematic utterances (Bailey, 2006; Fulcher, 2003). Its assessment is uniquely challenging because it requires teachers to make instantaneous judgments on multiple linguistic strands simultaneously (Luoma, 2004). This real-time processing creates a significant burden for assessors who must analyze student production at the moment of delivery (Inayah et al., 2019).

Furthermore, the evaluative criteria used in speaking assessment have historically reflected a traditional skills-and-elements model. Lado (1961) and Carroll (1961), as cited in Bachman and Palmer (2010), conceptualized language ability in terms of discrete components, including grammar, vocabulary, and phonology, assessed through the four language skills. This model has historically influenced both language teaching and testing. Recent empirical evidence continues to prioritize these discrete microlinguistic features while neglecting discourse-level competencies like interactional meaning negotiation and pragmatics (Tajeddin et al., 2018; Wang, 2017). This narrow conceptualization is often reinforced by a "testing as you were tested" mentality, where teachers replicate the teacher-centered, rehearsed assessment formats they experienced as learners (Vogt & Tzagari,

2014). While there is an emerging awareness of authentic assessment for speaking, teachers often struggle to balance formative feedback with the summative requirement of assigning final grades (Derakhshan & Ghiasvand, 2022; Inayah et al., 2019; Wang, 2017).

Novice EFL Teachers and Assessment Practices in Indonesia

The development of assessment literacy among novice teachers represents an important transitional phase as they move from professional preparation into classroom practice. In Indonesia, *PPG Prajabatan* is a professional teacher education program for prospective teachers following undergraduate study and leading to teacher certification (Direktorat Jenderal Guru dan Tenaga Kependidikan, 2023). Assessment is explicitly included within the program framework, with the 2023 technical guidelines identifying assessment as part of the program's scope and the current *PPG* curriculum including *Prinsip Pengajaran dan Asesmen I dan II* or Principles of Teaching and Assessment I and II as core coursework (Direktorat Jenderal Guru dan Tenaga Kependidikan, 2023; Direktorat Jenderal Guru, Tenaga Kependidikan, dan Pendidikan Guru, 2024; *PPG Bagi Calon Guru - Direktorat Pendidikan Profesi Guru (PPG)*, n.d.). The inclusion of assessment within the *PPG* framework provides relevant professional context for examining how novice teachers demonstrate and enact assessment knowledge once they enter their own classroom contexts. Existing research, however, suggests that assessment knowledge does not always translate straightforwardly into classroom assessment practice. Damayanti et al. (2024), for example, found that Indonesian secondary EFL teachers reported strong assessment knowledge and positive perspectives on assessment principles, while classroom observations indicated more limited implementation of assessment practices.

Recent studies on pre-service and novice teachers in Indonesia indicate that satisfaction with university assessment courses does not necessarily equate to assessment preparedness in the field (Farmasari et al., 2023). In addition, while the teaching practicum helps elevate literacy from insufficient to a moderate level, novice teachers still struggle with technical skills like rubric construction and the use of statistics for establishing reliability (Prastikawati et al., 2024). Within the TALiP framework, such tensions can be understood in terms of how teachers negotiate competing demands and contextual constraints when making assessment decisions, such as institutional accountability mandates like the Minimum Completion Criteria (KKM) and heavy administrative workloads (Xu & Brown, 2016). Consequently, novice teachers in Indonesia often rely on senior colleagues or MGMP (teacher support group) guidelines rather than exercising autonomous professional judgment (Luthfiyyah et al., 2020a)

While previous research has contributed to our understanding of language assessment literacy among EFL teachers, several areas remain underexplored. First, research specifically addressing classroom-based assessment literacy for speaking remains

relatively limited compared with broader studies of language assessment literacy, although Tajeddin et al. (2018) have directly examined novice and experienced EFL teachers' speaking assessment knowledge and practices. Second, Indonesian studies have examined assessment literacy through a range of methods, including questionnaires, interviews, and, in some cases, classroom observation (Aria et al., 2021; Damayanti et al., 2024; Fitriyah et al., 2022; Zulaiha et al., 2020). However, fewer studies have combined scenario-based elicitation of assessment knowledge with direct observation of classroom assessment and analysis of instructional documents, particularly in relation to speaking assessment. Third, less is known about how novice English teachers who have completed Indonesia's *PPG Prajabatan* program enact classroom-based speaking assessment in their actual classroom contexts. This study addresses these areas by examining assessment knowledge demonstrated through scenario responses and its relationship to classroom implementation, drawing on interviews, classroom observations, and instructional documents.

METHOD

Research Design

This study employed a qualitative multiple-case study design (Yin, 2018) to examine classroom-based speaking assessment literacy among novice English teachers who had completed *PPG Prajabatan*. In accordance with case study methodology in applied linguistics, each case was bounded as an individual novice EFL teacher situated within the institutional and instructional context of their workplace (Duff, 2018). The study comprised two phases. Phase 1 examined participants' assessment knowledge through scenario-based tasks and semi-structured interviews. Phase 2 focused on the classroom enactment of speaking assessment through classroom observations, follow-up interviews, and analysis of instructional documents involving three focal cases.

Participants and Context

Participants were recruited in West Java through convenience and snowball sampling. To be eligible, teachers had to have completed *PPG Prajabatan*, have no more than five years of teaching experience, and be currently teaching English. Although *PPG Prajabatan* is a pre-service teacher professional education pathway, all participants had gained teaching experience before completing *PPG Prajabatan*. Thus, *PPG Prajabatan* completion did not mark the beginning of their teaching experience. Ten teachers who met these criteria volunteered to participate.

The researcher was a university lecturer who was not involved as a *PPG* educator and had prior professional acquaintance with three of the ten participants. These participants helped circulate the study invitation to other *PPG* graduates in their professional networks,

contributing to the snowball recruitment process. The three focal teachers who subsequently participated in the classroom observation phase (P1, P9, and P10) were not among those previously known professionally to the researcher.

The study did not receive formal approval from an institutional research ethics committee. Nevertheless, consent and permission procedures were implemented for the different groups involved in the study. The ten participating teachers provided electronic informed consent through the Google Form accompanying the scenario-based task. For the classroom observations, permission was obtained from the school leader and from the participating teachers. Before each observation, the researcher explained to the students that the researcher would be present in the classroom and that classroom activities would be audio-recorded for research purposes. The students were invited to provide verbal assent before the observation began. No separate parental or guardian consent was obtained. Participants' names were replaced with codes (P1–P10) in the analysis and reporting.

All ten participants happened to have completed *PPG Prajabatan* at the same *PPG* institution in West Java, with completion years ranging from 2023 to 2025. Although the *PPG Prajabatan* completion year was not imposed as a separate sampling criterion, this range reflects the completion years of the ten teachers who volunteered and met the eligibility criteria. The participants taught at primary, junior high, senior high, and vocational schools, and reported between less than one year and three years of teaching experience at the time of data collection. The term novice English teachers in this study therefore refers to teachers with no more than five years of teaching experience who had completed *PPG Prajabatan* and were currently teaching English. Their characteristics and participation across the data-collection stages are presented in Table 1.

Table 1

Participant characteristics and participation across data-collection stages

Partici pant	Teachi ng Experi ence	Gen der	Teaching Context	Acade mic Qualifi cation	<i>PPG</i> Pathw ay	<i>PPG</i> Comple tion Year	Scen ario	Initi al Inter view	Classr oom Obser vation
P1	3 years	Fem ale	Junior High School	B.Ed.	Prajab atan	2023	✓	✓	✓
P2	1 year	Fem ale	Junior High School	B.Ed.	Prajab atan	2023	✓	✓	

P3	2 years	Male	Junior High School	B.Ed.	Prajabatan	2025	✓	✓	
P4	1 year	Female	Primary School	B.Ed.	Prajabatan	2025	✓	✓	
P5	<1 year	Female	Junior High School	B.Ed.	Prajabatan	2025	✓		
P6	3 years	Female	Vocational High School	B.Ed.	Prajabatan	2025	✓		
P7	1/6 year	Female	Senior High School	B.A.	Prajabatan	2025	✓		
P8	1 year	Female	Primary School	B.A.	Prajabatan	2025	✓	✓	
P9	1 year	Female	Vocational High School	B.A.	Prajabatan	2023	✓	✓	✓
P10	1 year	Female	Vocational High School	B.Ed.	Prajabatan	2023	✓	✓	✓

All ten participants completed the scenario-based task. Seven subsequently participated in individual semi-structured interviews, and three of these seven participants—P1, P9, and P10—volunteered and were available to participate in classroom observations, becoming the focal cases for Phase 2. Each focal teacher was observed in two instructional sessions, resulting in six observation sessions in total. P1 was observed in the same Grade 9 class across two sessions, whereas P9 and P10 were each observed in two different classes, with the same topic taught in each teacher's two observed classes. The three focal teachers also participated in post-observation follow-up interviews and provided relevant instructional documents. The characteristics of the three focal classroom contexts are presented in Table 2.

Table 2

Characteristics of the three focal classroom cases

Characteristic	Case 1: P1	Case 2: P9	Case 3: P10
School context	Junior High School	Vocational High School	Vocational High School
Grade	9	10	11

Class size	24	24	16
Classes observed	Same class	2 different classes	2 different classes
Topic	<i>Holiday Experience</i>	<i>Giving Directions</i>	<i>Giving Opinions</i>
Observed speaking task	Group presentation	Pair dialogue performed for the class	Pair dialogue performed for the class
Observation duration	2 × 2 × 40 min	2 × 2 × 45 min	2 × 2 × 45 min
Student proficiency	No standardized measure; the teacher described learners as having basic English proficiency	No standardized measure; the teacher described learners as having basic English proficiency	No standardized measure; the teacher described learners as having basic English proficiency
Assessment role	Contributed to semester results; exact weighting not documented	Contributed to semester results; exact weighting not documented	Contributed to semester results; exact weighting not documented

Data Collection

Scenario-Based Tasks

All ten participants completed a scenario-based task through Google Forms. The task consisted of three scenarios presenting classroom-based speaking assessment situations. Participants were asked to make and explain assessment decisions concerning assessment purposes, content/criteria, and assessment methods, corresponding to the three dimensions of Xu and Brown's (2016) TALiP framework. The scenarios were presented in Indonesian, and participants responded in the same language. They were given approximately 25 to 30 minutes to complete the task. The scenarios were developed by the researcher based on the study's research questions and the three TALiP dimensions and were contextualized to classroom-based speaking assessment in Indonesian EFL settings, and adapted from scenario-based language assessment literature (e.g., Tajeddin et al., 2018). Scenario-based methods (or vignettes) can be used to elicit teachers' beliefs, knowledge, and decision-making by presenting them with contextualized situations that require interpretation and response (Möller et al., 2025; Stravakou & Lozgka, 2018; Tajeddin et al., 2018).

Semi-Structured Interviews

Seven of the ten participants subsequently participated in individual face-to-face semi-structured interviews lasting approximately 30 minutes. The interviews focused on participants' reasoning underlying their scenario responses and included follow-up questions to clarify and elaborate their written responses, particularly in relation to

assessment purposes, content/criteria, and methods. The interviews were conducted in Indonesian and audio-recorded with participants' permission. The researcher transcribed the recordings for analysis. Relevant excerpts reported in English were translated from the Indonesian transcripts by the researcher and checked against the original transcripts to preserve the intended meaning.

Classroom Observations

Classroom observations were conducted with the three focal teachers (P1, P9, and P10). The researcher personally conducted all observations as a non-participant observer. Six instructional sessions were observed in total. P1 was observed twice in the same Grade 9 class, while P9 and P10 were each observed in two different Grade 10 and Grade 11 classes, respectively. P1's two observations addressed the topic *Holiday Experience*: the first focused on teacher modelling of the text through reading activities and grammar explanation, while the second focused on students' group presentations of their work. P9 taught *Giving Directions* in each of the two observed classes, with the topic introduced and completed within each session. Similarly, P10 taught *Giving Opinions* in each of the two observed classes, with the topic introduced and completed within each session. Each observation comprised two 40-minute periods for P1 and two 45-minute periods for P9 and P10.

Field notes were guided by the study's analytical focus on three dimensions of classroom-based speaking assessment: assessment purposes, assessment content/criteria, and assessment methods. During each observation, the researcher documented the sequence of classroom activities and noted how speaking assessment was enacted, including the purposes of assessment, the aspects of speaking attended to, the criteria evident in classroom activities, and the assessment methods used. Classroom interaction was audio-recorded to support the field notes during analysis.

Instructional Documents and Follow-Up Interviews

Relevant instructional documents supplied by the three focal teachers, including lesson plans, syllabi, and available scoring rubrics, were collected to provide additional evidence for interpreting classroom assessment practices. Following the observations, the three focal teachers participated in follow-up interviews to clarify assessment practices and decisions identified during the observed lessons.

Data Analysis

The data collected in Phase 1 were analyzed to address RQ1 concerning participants' assessment knowledge for classroom-based speaking assessment. The scenario-based task responses were analyzed using directed content analysis (Hsieh & Shannon, 2005). The TALiP framework (Xu & Brown, 2016) provided the initial deductive analytical structure,

comprising three dimensions: assessment purposes, assessment content/criteria, and assessment methods. The specific categories within these dimensions were then developed from the assessment ideas expressed in participants' responses. Because participants could articulate more than one purpose, criterion, or method, the categories were not mutually exclusive. The researcher was the sole coder and recorded the coding in a matrix using Microsoft Excel. Emerging categories and interpretations were discussed with two other members of the research team to clarify ambiguous cases and refine interpretations where necessary. Because the other researchers did not independently code the dataset, no inter-coder agreement was calculated.

Descriptive frequency counts were calculated to summarize how many participants articulated each category in their scenario responses. These counts were used to describe patterns in the assessment ideas expressed by participants, rather than to quantify the completeness or depth of their overall assessment knowledge. Accordingly, the findings are interpreted as assessment knowledge articulated in response to the scenarios. The seven initial interviews were analyzed using the same three TALiP dimensions and the categories developed from the scenario responses. Interview data were used to clarify and elaborate on participants' reasoning underlying their scenario responses.

The Phase 2 data were analyzed to address RQ2 concerning how the assessment knowledge demonstrated by the three focal teachers related to their classroom implementation of speaking assessment. The classroom observation data, instructional documents, and follow-up interviews were analyzed using content analysis, with the three TALiP dimensions—assessment purposes, content/criteria, and methods—serving as the initial deductive coding framework (Xu & Brown, 2016). The analysis involved reviewing the observation field notes and revisiting the accompanying audio recordings to verify and clarify relevant classroom interactions, identifying assessment-related practices, and coding them according to the three dimensions. The audio recordings were used to support the field notes rather than as a separately transcribed dataset. Instructional documents were examined to corroborate observed practices, while follow-up interviews were analyzed to clarify teachers' assessment decisions and intentions. Claims about the presence or absence of particular assessment practices were verified by cross-checking the observation records, audio-supported evidence, instructional documents, and relevant follow-up interview responses.

Trustworthiness was addressed through credibility, dependability, confirmability, and transferability. Credibility was supported through triangulation across scenario responses, interviews, classroom observations, instructional documents, and follow-up interviews. Member checking was also conducted through WhatsApp communication with participants, in which the researcher shared relevant interpretations of their assessment

practices and invited participants to clarify or correct the interpretations where necessary. Participants' responses were considered in refining the interpretations where clarification was provided. Dependability was supported through an audit trail comprising the coding matrices, records of category development, and documented analytical decisions throughout the analysis. Confirmability was addressed through explicit researcher positionality and by grounding interpretations in evidence from multiple data sources. Transferability was supported through detailed descriptions of the three focal classroom contexts, including school level, grade, class size, instructional topic, observed activities, and session duration (Table 2).

RESULTS AND DISCUSSION

The findings are reported in two parts corresponding to the research questions. The first describes the ten teachers' knowledge of speaking-assessment purposes, criteria, and methods, drawing primarily on the scenarios and follow-up interviews. The second examines how CBA was enacted in the three observed classrooms, with observation records considered alongside instructional documents and follow-up interviews.

Results

Novice English teachers' knowledge of classroom-based speaking assessment

The findings for RQ1 are presented across the three TALiP dimensions of assessment purposes, content/criteria, and methods. Table 3 summarizes the assessment categories articulated by the ten participants in their scenario responses, together with the number of participants articulating each category and illustrative excerpts.

Table 3

Assessment categories demonstrated in the scenario responses

TALiP dimension	Assessment category	Participants articulating the category (n=10)	Illustrative excerpt
Assessment purposes	Formative	8	"I conduct formative assessment to find out the extent of students' abilities throughout the learning process." — P3
	Summative	6	"The purpose of speaking assessment is to find out how far students can pronounce or produce English, and even use it." — P4
Assessment content/criteria	Pronunciation	8	"The aspects I will use include pronunciation, fluency, confidence, and content." — P1
	Fluency	8	"The aspects I will use include pronunciation, fluency, confidence, and content." — P1

	Vocabulary	7	"In the rubric, I usually assess pronunciation, fluency, vocabulary, grammar, as well as content and confidence when speaking." — P5
	Grammar	5	"The aspects I will include in the rubric are pronunciation, fluency, vocabulary, grammar, content, and confidence." — P4
	Content	6	"I will adjust [the criteria] to the learning objectives to be achieved, such as speaking fluency, pronunciation, and content delivery, while also considering the students' abilities." — P1
	Organization	1	"Organization (structure): how students organize the opening, body, and closing so that the presentation has a clear flow." — P8
	Confidence	7	"The criteria I use in speaking activities are appropriate pronunciation, fluency, and, most importantly, confidence." — P7
Assessment methods	Presentation/monologue	8	"The speaking assessment method I most frequently use in class is presentation, both group and individual." — P10
	Role-play	4	"For example, when teaching descriptive text, I assess students through role-play. They are given cards that they need to act out, allowing them to speak spontaneously." — P3
	Dialogue	2	"I use rubric-based assessment with presentation and dialogue tasks as the main techniques." — P1
	Question-and-answer	1	"Impromptu. I provide clues/rules about what students need to present, and then they are asked to respond directly." — P7

Novice English teachers' knowledge of speaking assessment purposes

Formative purposes were reported by eight of the ten teachers in their scenario responses. They described speaking assessment in terms of supporting learning, for example by monitoring progress, giving feedback, or reinforcing what students had learned. Summative purposes were also articulated by six of the ten teachers, who referred to evaluating students' speaking achievement or ability. The responses therefore did not separate formative and summative assessment into mutually exclusive functions. Rather, several teachers articulated speaking assessment as serving learning while also contributing to judgments about achievement. These findings represent the assessment

knowledge articulated in response to the scenarios and should not be interpreted as indicating the presence or absence of participants' overall assessment knowledge.

The interviews helped explain this pattern. P2 described speaking assessment as *"more focused on helping students improve their speaking, not just on giving scores,"* while P5 emphasized the opportunity to *"give immediate feedback and see how far students understand the material."* These comments were consistent with the formative orientation found in the scenarios. At the same time, P4 noted the need to *"see students' final ability, even if it is not always included in midterm or final exam scores."* Thus, participants' accounts included both support for ongoing learning and evaluation of outcomes.

Novice English teachers' knowledge of speaking assessment criteria

The criteria articulated in the scenario responses were concentrated mainly on linguistic features. Pronunciation and fluency were each articulated by eight teachers, vocabulary by seven, and grammar by five. Six teachers referred to content, whereas only one mentioned organization. Confidence was also articulated by seven teachers. The responses included both linguistic and affective aspects of speaking performance.

Interview responses shed light on these choices. P1 explained that *"pronunciation and fluency are the most visible, so they are easier to assess when students speak,"* and P6 stressed whether students could convey meaning fluently and be understood. Grammar received less emphasis partly because teachers were concerned that close attention to errors might discourage students from speaking; P3, for example, said that grammar was not emphasized too much so that students would not be afraid of making mistakes. Organization was mentioned by only one participant in the scenario responses. One possible interpretation is that the limited attention to organization may be related to the teachers' focus on shorter utterances or brief responses rather than extended discourse. Confidence, by contrast, was articulated by several participants when describing speaking performance. As P5 observed, some students could speak but were *"not brave enough."*

Novice English teachers' knowledge of speaking assessment methods

Presentation or monologic performance was articulated by eight teachers, making it the most frequently mentioned assessment method. Role-play appeared in four responses, dialogue in two, and question-and-answer activities in one. Thus, individual performance tasks were articulated more often than methods requiring sustained two-way interaction.

The interviews suggest that practicality influenced these preferences. Participants described presentations and monologues as easier to organize and score, particularly in large classes. P2 noted that *"monologs are easier to assess because only one student performs at a time,"* while P4 considered them more time-efficient than dialogues or role-plays. P7

acknowledged the communicative value of role-play but noted that using it required more preparation and took up more classroom time. Teachers also associated spontaneous interaction with difficulties in maintaining assessment control and fairness across students with different proficiency levels. P3 summarized this concern by saying that *"spontaneous dialogue was difficult to assess individually within the available time."*

Novice English teachers' classroom enactment of speaking assessment

The three focal cases differed in school context, grade level, class size, instructional topic, speaking task, and observation duration (Table 2). P1 taught Grade 9 students in a junior high school, P9 taught Grade 10 students in a vocational high school, and P10 taught Grade 11 students in a vocational high school. P1's observations involved a group presentation on *Holiday Experience*, whereas P9 and P10 conducted pair-dialogue activities on *Giving Directions* and *Giving Opinions*, respectively. The observed classes ranged from 16 to 24 students, and each teacher was observed in two sessions. The teachers described their students as having basic English proficiency, although no standardized proficiency measure was available. In all three cases, the speaking assessment contributed to students' semester results, although the exact weighting was not documented.

Despite these contextual differences, the three observed cases showed a broadly similar instructional sequence. Lessons began with teacher questioning and presentation of objectives, followed by input through video or topic introduction and a language-focus stage involving explanation and translation. Students then prepared a dialogue or group oral presentation, performed it, received feedback, and moved to the closing stage. Speaking assessment occurred within this instructional sequence rather than as a clearly separated event.

Novice English teachers' speaking assessment purposes in classroom practice

Across the three cases, speaking assessment served both learning and grading purposes. The observations showed three recurring features: assessment embedded in regular instructional activities, teacher monitoring and feedback during or after performance, and the recording of performance scores.

Assessment embedded in instructional activities

None of the three teachers introduced speaking assessment as a separate, explicitly labeled stage of the lesson. Instead, assessment occurred through activities such as pair dialogues and group presentations. Field notes, for example, recorded students constructing a dialogue before performing it for the class and groups presenting a talk about a holiday experience. Although these were introduced as classroom tasks, teachers observed the performances and used them as opportunities to evaluate students' speaking.

Formative purposes

A formative function was visible in the feedback given around students' performances. Teachers corrected pronunciation, commented on clarity, or provided broader feedback after a task. In one lesson, a teacher corrected a student's pronunciation of "five" during the dialogue and asked the student to repeat the word. In another, feedback followed each pair's performance; elsewhere, comments were given to the class after group presentations. These episodes show that assessment information was sometimes used immediately to respond to students' performance.

Summative purposes

The same performances also contributed to grading. In each classroom, the teacher was observed writing notes while students performed. Follow-up interviews clarified that these notes included performance scores recorded in grade books. The teachers explained that the scores were used as part of students' semester results. Thus, the classroom performances had a dual role: they generated opportunities for feedback while also contributing to formal evaluation.

Novice English teachers' speaking assessment criteria in classroom practice

Two patterns were especially evident in the assessment criteria used in practice. First, criteria were largely controlled by the teacher and were not made explicit to students during the observed tasks. Second, the content and timing of teachers' feedback provided indirect evidence of the aspects of speaking to which they were attending.

Implicit and teacher-centered assessment criteria

The observations showed a clear distinction between planned criteria and their classroom communication. All three teachers evaluated student performances and recorded scores, and their lesson plans contained assessment criteria. During the observed lessons, however, the criteria were not explicitly communicated to students when the speaking tasks were assigned. Field notes from all six observed sessions recorded no explicit explanation of the speaking-assessment criteria before students performed.

The documents presented a different picture at the planning level: lesson plans listed criteria such as pronunciation, vocabulary, fluency, accuracy, and confidence. The issue, therefore, was not the complete absence of criteria but the lack of observed communication of those criteria during classroom implementation. There was therefore no observed evidence that students were explicitly informed of the criteria used to assess their speaking performance during the observed tasks.

The follow-up interviews provided information about the teachers' assessment practices and decisions, but they did not provide direct evidence about the teachers' reasons for communicating or not communicating the criteria before students' performances. Nor

did the available data establish whether the teachers considered sharing the criteria through other means. Thus, the available evidence supports a description of a planning–practice discrepancy rather than an explanation of the reasons underlying the observed lack of criterion communication.

Range and variation of feedback

Feedback differed across the three classrooms in both timing and focus, reflecting different ways of responding to students' speaking performance.

P9, in the first observed session, responded to students' performances while they were performing their dialogue. When students produced directions such as "turn right," the teacher interrupted by asking where they should turn. The field notes indicate that the teacher provided immediate feedback to each pair and also gave oral feedback concerning comprehension. In the second observed session, feedback was delivered more broadly after the performances. The teacher reminded students, "*Kalau presentasi selanjutnya suaranya yang kencang, dan perhatikan teman yang sedang presentasi*" (For the next presentation, speak louder and pay attention to the classmate who is presenting). Thus, P9's feedback varied from immediate prompts concerning students' utterances and comprehensibility to broader comments concerning delivery and attention to the audience.

P10, in the first observed session, provided oral feedback after each pair performed. The field notes recorded comments such as "*Kalian sudah sangat percaya diri dan jelas artikulasinya*" (You were very confident and your articulation was clear), "*Next time jangan terlihat terlalu menghafal*" (Next time, try not to look as though you are memorizing), and "*Bukan menghafal tapi harus memahami apa yang kalian sampaikan*" (Rather than memorizing, you should understand what you are saying). The teacher subsequently addressed the whole class, emphasizing understanding rather than memorization and encouraging students to speak without fear of making mistakes. In the second observed session, the teacher again provided oral feedback immediately following each pair's performance, including "*Sangat pede, suara kurang jelas. Untuk meningkatkan harus terus belajar*" (Very confident, but your voice was not clear; to improve, you need to keep practising). The teacher's whole-class feedback also referred to pronunciation, confidence, and students' willingness to speak despite errors.

In the second observed session, P1's students presented group talks about a holiday experience using posters. The field notes indicate that most students read their prepared texts rather than delivering the talk spontaneously. After each presentation, P1 asked the class to give applause to the presenting group, but neither the observation notes nor the audio recording contains substantive verbal feedback on students' speaking performance.

In the follow-up interview, P1 explained that the activity was focused on performance and that she normally provided individual written feedback during independent practice. She explained that individual written feedback allowed her to identify what each student did well and what needed improvement, while avoiding the time required for immediate feedback after each presentation. She described the feedback as focusing particularly on pronunciation and confidence, with suggestions for further English-language practice to improve pronunciation.

Taken together, the three cases showed variation in both the timing and focus of feedback, reflecting different ways of responding to students' speaking performance. P9 combined immediate responses during students' performances with broader comments after the activity, addressing comprehensibility, delivery, and attention to the audience. P10 provided feedback immediately after performances, focusing on articulation, voice clarity, pronunciation, confidence, and students' reliance on memorization. In P1's case, the observed group presentation was treated primarily as a performance activity, while her follow-up account indicated that more individualized feedback was normally provided in written form during independent practice. Thus, feedback was not limited to a single mode of delivery: it ranged from immediate oral responses during or after performance to broader whole-class comments and individualized written feedback.

Novice English teachers' speaking assessment methods in classroom practice

Speaking was assessed mainly through oral performance tasks, but the observed implementation showed little variation in how those tasks were prepared and evaluated. Three patterns were found across the cases: reliance on scripted or memorized performance, limited instructional scaffolding before the assessment task, and little use of assessment methods that involved students in judging performance.

Dependence on memorization and scripted performance

The three classrooms used pair dialogues or group presentations as the main speaking-assessment tasks. One class prepared dialogues for asking and giving directions from a map; another used paired discussions that were later performed for the class; and the third used group presentations about holiday experiences supported by posters.

Preparation for these performances frequently involved scripts and memorization. Field notes recorded students memorizing dialogues after the teacher had checked the scripts, while most groups in the presentation lesson read their prepared text rather than speaking more spontaneously. The follow-up interviews showed that the teachers regarded this preparation as necessary for their learners. All three referred to students' basic English proficiency, and they doubted that students could complete the tasks without substantial

rehearsal. In these classrooms, memorization was therefore not simply a student strategy; it was accepted by teachers as a practical way of making the speaking task manageable.

Instructional preparation for assessment tasks

Before the performance tasks, the observed lessons included limited controlled speaking practice. After input and teacher explanation, students generally moved to constructing a dialogue or monologue. The lesson plans showed the same sequence and did not include a separate controlled-practice stage. Field notes also showed that preparation centered on having scripts checked and memorizing them. These observations describe how the speaking tasks were prepared instructionally and indicate that performance practice was closely integrated with task preparation rather than organized as a separate controlled-practice stage.

Teacher-led assessment methods

Assessment methods were primarily teacher-led across the three observed classrooms. No peer- or self-assessment activity was observed, and the lesson plans did not provide rubrics or procedures for students to assess their own or their peers' speaking. Assessment decisions therefore remained with the teacher. One classroom episode illustrates this teacher-centered pattern. After a pair performed, the teacher asked the class what the two students had been expressing opinions about. When no student responded, the teacher answered the question herself.

Discussion

The findings show that the relationship between articulated assessment knowledge and classroom implementation varied across the three TALiP dimensions of assessment purposes, content/criteria, and methods. The scenario responses from all ten participants, supplemented by interviews with seven participants showed that speaking assessment was understood as serving both formative and summative purposes, and that participants recognized a range of linguistic and affective dimensions of speaking performance. In the three focal classrooms, speaking assessment likewise served both learning and grading purposes, while assessment criteria were present in planning documents but were not explicitly communicated during the observed speaking tasks. The observed assessment methods involved oral performance tasks, consistent with methods articulated in the knowledge data, but their implementation was shaped by contextual considerations such as perceived learner proficiency, time, and task manageability. Thus, the findings do not indicate a uniform transfer or failure to transfer assessment knowledge into practice; rather, the relationship varied by assessment domain and by the conditions under which assessment was enacted.

Regarding assessment purposes, the prominence of formative purposes in the scenario responses indicates that participants articulated speaking assessment as having a role in supporting learning. Eight of the ten participants referred to formative purposes, describing assessment in terms of monitoring students' progress, providing feedback, or supporting learning. At the same time, six participants also articulated summative purposes, indicating that speaking assessment could serve both learning and achievement-related functions. This combination is consistent with the dual role of classroom assessment, in which assessment may be integrated into instruction while also contributing to judgments of achievement (Damayanti et al., 2024; Poehner, 2009; Wang, 2017). In the three observed classrooms, speaking assessment likewise served both purposes: teachers responded to students' performances through feedback, while the performances also contributed to students' semester results. The findings therefore show correspondence between the formative and summative purposes articulated in the knowledge data and the dual purposes evident in the three observed cases, although the available data do not establish the extent to which the teachers' practices reflected a broader assessment-for-learning framework.

The criteria identified by the teachers also reveal a particular view of speaking ability. Pronunciation, fluency, and vocabulary dominated the scenario responses, whereas discourse organization received little attention. These features are important, but a strong emphasis on discrete linguistic components resembles the traditional skills-and-elements view of language ability (Lado, 1961 & Carroll, 1961, as cited in Bachman & Palmer, 2010; Purpura, 2016). Empirically, this microlinguistic bias mirrors findings among secondary EFL teachers who prioritize immediate, observable aspects of speech over complex discourse competencies (Tajeddin et al., 2018; Wang, 2017). A plausible explanation for this practice is that such features are comparatively visible during short classroom performances, while speaking assessment requires teachers to attend to several aspects of performance at once Luoma (2004).

Furthermore, the frequent inclusion of confidence as an assessment criterion across scenario responses (n=7) and planning documents highlights a critical construct-validity issue in classroom speaking assessment. Similar attention to affective and non-achievement features has been documented in Indonesian EFL assessment research (Damayanti et al., 2024; Inayah et al., 2019; Prastikawati et al., 2024). While several participants treated confidence as an important indicator of speaking performance, measurement theory distinguishes between target achievement constructs and non-cognitive enabling factors such as effort, attitude, and psychological readiness (Brookhart, 2013; McMillan, 2011). Although self-confidence influences a learner's willingness to communicate in classroom settings (Inayah et al., 2019; Xu & Brown, 2016), its relevance to classroom participation

does not by itself establish that confidence should be incorporated into scores of English oral proficiency. If confidence is incorporated directly into scoring when the intended construct is English oral proficiency, doing so may introduce construct-irrelevant variance (Bachman & Palmer, 2010; Fulcher, 2012), potentially affecting the interpretation of scores as measures of oral proficiency. The finding therefore highlights an assessment-literacy issue concerning the need to distinguish between characteristics that may facilitate classroom speaking performance and the abilities that an assessment is intended to measure.

A further issue concerns the relationship between planned criteria and their classroom implementation. While the lesson plans contained assessment criteria, the criteria were not explicitly communicated to students during the observed speaking tasks. Making assessment criteria accessible to learners can support their understanding of learning targets and performance expectations and contribute to self-regulation and goal-directed learning (Brookhart, 2013; Li & Gu, 2023). By leaving criteria implicit and teacher-controlled during the observed tasks, the practice limited one potential formative function of assessment: making performance expectations accessible to learners. This maintained a predominantly teacher-controlled assessment process, even though the participants had articulated formative purposes for speaking assessment. Such a distinction is consistent with research showing that teachers may recognize formative purposes while classroom assessment remains strongly teacher-controlled (Aria et al., 2021; Fitriyah et al., 2022; Xu & Brown, 2016). However, the available data do not establish why the criteria were not communicated or whether they were shared through other channels. The finding is therefore best understood as a planning–practice discrepancy rather than direct evidence of inadequate assessment literacy.

A further issue concerns the relationship between assessment methods and the conditions under which speaking tasks were implemented. In the scenario responses, presentation/monologue was the most frequently articulated assessment method, appearing more often than role-play, dialogue, or question-and-answer formats. The interview data indicate that practicality influenced these preferences: participants described presentations and monologues as easier to organize and score in large classes and more time-efficient than interactive formats. They also associated spontaneous interaction with difficulties in maintaining assessment control and fairness across students with different proficiency levels; as P3 explained, spontaneous dialogue was difficult to assess individually within the available time. These concerns are consistent with the practical challenges of speaking assessment, which requires teachers to make judgments about multiple aspects of oral performance during real-time interaction (Inayah et al., 2019; Luoma, 2004).

These considerations are important for interpreting the observed reliance on scripted and rehearsed performance. In the three focal classrooms, the teachers described their learners as having basic English proficiency and regarded substantial preparation as necessary for students to complete the speaking tasks. Rehearsal therefore functioned as a practical accommodation rather than an inherently inappropriate assessment practice, helping make the speaking tasks achievable for learners while also making assessment more manageable within the teachers' classroom conditions. Similar tensions between communicative assessment goals and practical classroom constraints have been documented in EFL speaking-assessment contexts (Norhasim & Mohamad, 2020; Tajeddin et al., 2018).

At the same time, the accommodation had implications for the evidence generated by the assessment task. When students relied extensively on prepared or memorized language, teachers had fewer opportunities to observe spontaneous language production and interaction. This is important because speaking assessment involves more than the accurate production of prepared linguistic forms; it also concerns the ability to use language to communicate meaning in oral performance (Fulcher, 2003; Luoma, 2004). The assessment method therefore involved a trade-off between making performance feasible and manageable under classroom constraints and generating broader evidence of students' spontaneous oral language use. This pattern illustrates the contextual nature of assessment literacy emphasized in the TALiP framework, in which teachers' assessment decisions are shaped by the instructional and institutional conditions in which assessment is enacted (Xu & Brown, 2016).

Feedback practices in the observed classrooms further illustrate how assessment was enacted. Teachers responded to student performances through immediate oral feedback on features such as pronunciation and voice clarity, as well as post-performance comments concerning preparation, delivery, confidence, and reliance on scripts (P9, P10). For P1, the follow-up interview also indicated that the teacher normally provided individual written feedback during independent practice, particularly concerning pronunciation and confidence. Across the three focal classrooms, feedback remained predominantly teacher-led, and no peer- or self-assessment activity was observed. Because peer- and self-assessment also did not emerge among the assessment methods articulated in the scenario responses, the classroom evidence does not indicate a discrepancy between teachers' stated knowledge and observed practice in this respect. Rather, the correspondence between the two data sources suggests that teacher-led assessment was a consistent feature of the assessment repertoire represented in this study.

The absence of structured student involvement is nevertheless relevant from a learning-oriented assessment perspective. Peer- and self-assessment can provide learners

with opportunities to evaluate performance, engage with assessment criteria, and develop evaluative and self-regulatory capacities (Carless, 2007; Chen, 2008; Shepard, 2019). Research on LOA has also identified contextual factors such as time constraints, large classes, and teachers' limited knowledge or training as challenges to implementing more learner-involving assessment practices (Derakhshan & Ghiasvand, 2022). These factors may be relevant to understanding the feasibility of such practices in EFL classrooms, although the present study did not collect sufficient evidence to determine whether they influenced the teachers' assessment decisions. Thus, the absence of peer- or self-assessment in the observed lessons should not be interpreted as evidence of inadequate assessment literacy; rather, it indicates that student-led assessment was not part of the assessment practices articulated or observed in the cases examined.

The limited instructional preparation observed before the speaking tasks can also be considered in relation to the learning-oriented function of assessment. In the observed lessons, students generally moved from teacher explanation and script preparation to performance without a separate stage of controlled speaking practice. This finding is primarily an instructional-design feature rather than a direct indicator of assessment literacy. Its relevance to assessment lies in how the preparation sequence shaped the type of performance available for assessment. Because students moved relatively quickly toward scripted or rehearsed performance, the resulting tasks provided greater opportunity to assess prepared delivery than more spontaneous oral production. From a learning-oriented assessment perspective, assessment tasks are intended not only to generate evidence of performance but also to contribute to learning through appropriate task design, support, and feedback (Carless, 2015; Fulcher, 2021). The limited preparation observed in these cases therefore illustrates how instructional design and assessment method intersect: the way performance was prepared influenced the range of speaking evidence available for assessment. This points to a practical area for professional development, as novice teachers may benefit from opportunities to design assessment sequences in which modeling, guided practice, criteria, performance, and feedback are deliberately connected (Fitriyah et al., 2022; Li & Gu, 2023). At the same time, the teachers' reliance on rehearsal should be understood in context, as they described their learners as having basic English proficiency and regarded substantial preparation as necessary to make the tasks manageable.

The findings also suggest several tentative implications for early-career teacher development. First, the discrepancy between criteria documented in lesson plans and their lack of explicit communication during the observed tasks suggests that professional learning could provide opportunities for teachers to practice translating assessment criteria into learner-accessible expectations. Second, the reliance on scripted and rehearsed performance suggests the value of activities in which novice teachers examine how task

preparation can balance learners' proficiency and classroom feasibility with opportunities to generate broader evidence of oral performance. Third, the observed variation in feedback practices suggests that professional learning could support teachers in connecting assessment evidence and feedback with subsequent instruction. These implications are intended as areas for professional learning rather than recommendations for changes to the *PPG* program, since the present study did not examine *PPG* curriculum content, the amount of assessment practice provided during the programme, or participants' preparation for speaking assessment within *PPG*.

CONCLUSION

This study investigated classroom-based speaking assessment literacy among novice English teachers, examining the relationship between assessment knowledge articulated in scenario responses and classroom enactment in three focal cases. The findings show that this relationship varied across assessment domains: assessment purposes showed broad correspondence between articulated knowledge and classroom practice, assessment criteria revealed a planning–practice discrepancy, and assessment methods showed both correspondence and contextual adaptation.

The ten participants generally described speaking assessment in both formative and summative terms and identified linguistic as well as affective criteria. In the three observed classrooms, speaking assessment likewise served both learning and grading purposes. However, although assessment criteria were present in lesson plans, no explicit communication of these criteria to students was observed during the speaking tasks. Speaking performances also relied substantially on scripted and rehearsed language, which the teachers described as necessary for learners with basic English proficiency but which limited opportunities to observe more spontaneous oral language use. Assessment remained primarily teacher-led, and no peer- or self-assessment was observed. Taken together, these findings suggest that assessment knowledge was not uniformly transferred or failed to transfer into practice; rather, its relationship with classroom enactment differed according to the assessment domain and the contextual conditions in which assessment was implemented.

By focusing specifically on speaking, this study adds a more fine-grained account of how novice teachers negotiate the particular demands of assessing oral performance. While implicit criteria, teacher-led assessment, and limited student involvement are not unique to speaking assessment, the present findings identify several features that are particularly relevant to the assessment of speaking. Participants most frequently articulated presentation/monologue as an assessment method, while the observed classrooms relied substantially on scripted and rehearsed oral performances. The teachers described this

preparation as necessary for learners with basic English proficiency and compatible with the practical demands of classroom assessment. At the same time, extensive reliance on prepared language limited opportunities to observe spontaneous oral production and interaction. The study therefore extends understanding of novice teachers' speaking assessment by showing how contextual conditions such as learner proficiency, class size, available time, manageability, and perceived fairness can shape not only teachers' choice of speaking-assessment methods but also the type of oral performance evidence available for assessment. This highlights an important aspect of speaking assessment: teachers need to balance tasks that are practical and manageable with tasks that provide meaningful evidence of students' speaking ability.

This study has several limitations that should be considered when interpreting the findings. First, the participants were recruited through convenience and snowball sampling, and all ten participants happened to have completed *PPG Prajabatan* at the same institution in West Java. The participants were also volunteers, and the snowball process was initiated through three teachers who had prior professional acquaintance with the researcher. Although the three focal teachers observed in Phase 2 were not previously known to the researcher, the sampling process may still have favored teachers who were accessible or willing to participate. The findings should therefore not be generalized beyond the participant group examined in this study or taken as evidence about the *PPG* program as a whole.

Second, the depth of evidence differed across the two research questions. All ten participants contributed scenario responses for RQ1, seven participated in initial interviews, but only three volunteered and were available for classroom observations, follow-up interviews, and document analysis. The classroom findings therefore represent three focal cases rather than the broader group of ten participants. In addition, the observed teachers described their students as having basic English proficiency, but no standardized proficiency measure was used. This limits the extent to which the role of learner proficiency can be established beyond the teachers' own descriptions. The observations were also limited to two sessions per focal teacher, so the findings may not capture the full range of their assessment practices across topics, classes, or longer periods.

Third, the study did not collect data directly from students. Consequently, although no explicit communication of assessment criteria to students was observed during the focal speaking tasks, the study cannot determine whether students were aware of the criteria through other channels or how they understood the expectations for successful performance. Similarly, the study cannot establish how students perceived or used the feedback provided by their teachers. Future studies could incorporate student interviews, questionnaires, or student work to examine these dimensions more directly.

Finally, the researcher served as the sole coder of the data, although interpretations were discussed with two other research team members and supported through triangulation, member checking, and an audit trail. The researcher also had prior professional acquaintance with three of the ten participants, which may have influenced recruitment or interpretation despite efforts to maintain analytic reflexivity. These factors should be considered when interpreting the qualitative analysis. Future research could strengthen the evidence base through broader and more varied sampling, longer-term classroom observation, independent coding or additional analytic checks, and inclusion of student perspectives.

Future studies could follow a larger group of novice English teachers who have completed *PPG Prajabatan* over a longer period to examine how assessment knowledge and practice develop with experience. Including students' perspectives could also clarify how learners understand assessment criteria and feedback. Research across different school and regional contexts would help determine which of the patterns identified in these three cases recur elsewhere and which are specific to the settings examined here.

DECLARATIONS

Ethics Approval

This study did not receive formal approval from an institutional research ethics committee. Ethical procedures were nevertheless implemented throughout the study. Participation was voluntary, and informed consent was obtained electronically from all participating teachers before data collection. For classroom observations, permission was obtained from the participating schools and teachers. Students were informed about the researcher's presence and the audio recording of classroom activities and were invited to provide verbal assent before the observations. No separate parental or guardian consent was obtained.

Informed Consent

All participating teachers provided electronic informed consent before participating in the study. For classroom observations, participating teachers and school leaders provided permission for the observations and audio recording. Students present during the observed lessons were informed about the study and invited to provide verbal assent.

Data Availability

The data supporting the findings of this study are not publicly available because they contain potentially identifiable information from participating teachers and classroom observations.

Conflict of Interest

The authors declare that they have no conflict of interest related to this study.

Funding

This research was funded by the 2025 Internal Research Grant of Universitas Perjuangan Tasikmalaya (UNPER).

Author Contributions

Author 1 for Conceptualization, methodology, data collection, data analysis, and manuscript preparation. Author 2 for Contribution to data analysis and interpretation of the findings, and manuscript review. Author 3 for Contribution to data analysis and interpretation of the findings, and manuscript review. All authors reviewed and approved the final manuscript.

Declaration of AI and AI-Assisted Technologies

Generative AI (ChatGPT) was used during manuscript preparation to assist with language editing, clarity, and the revision of selected passages. The author critically reviewed and revised all AI-assisted suggestions and takes full responsibility for the accuracy, originality, interpretation, and final content of the manuscript.

Acknowledgements

The authors would like to sincerely thank the novice English teachers who participated in this study for generously sharing their perspectives, experiences, and classroom practices. We are also grateful to the participating schools and school leaders for their support and for facilitating access to the classroom settings in which the observations were conducted. Their time, cooperation, and openness were invaluable to the completion of this study.

REFERENCES

- Alderson, J. C., Brunfaut, T., & Harding, L. (2017). Bridging assessment and learning: A view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice*, 24(3), 379–387. <https://doi.org/10.1080/0969594X.2017.1331201>
- Anam, S., & Putri, N. V. W. (2021). How literate am I about assessment: Evidence from Indonesian EFL pre-service and in-service teachers. *English Review: Journal of English Education*, 9(2), 377–388. <https://doi.org/10.25134/erjee.v9i2.4374>
- Aria, D., Sukyadi, D., & Eri Kurniawan. (2021). Teacher assessment literacy: Indonesian EFL secondary teachers' self-perceived on classroom-based assessment practice. *English Review: Journal of English Education*, 10(1), 15–26. <https://doi.org/10.25134/erjee.v10i1.5349>
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Bailey, K. M. (2006). Issues in teaching speaking skills to adult ESOL learners. *Review of Adult Learning and Literacy*, 6, 113–164.
- Baker, B. (2016). Language assessment literacy as professional competence: The case of Canadian admissions decision makers. *Canadian Journal of Applied Linguistics*, 19(1), 63–83. <https://journals.lib.unb.ca/index.php/CJAL/article/view/23033>
- Brookhart, S. M. (2013). How to create and use rubrics for formative assessment and grading. *Ascd*.

- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57–66. <https://doi.org/10.1080/14703290601081332>
- Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, 69(6), 963–976. <https://doi.org/10.1007/s10734-014-9816-z>
- Chen, Y.-M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12(2), 235–262. <https://doi.org/10.1177/1362168807086293>
- Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, 71, 134–144. <https://doi.org/10.1016/j.tate.2017.12.010>
- Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43–56. <https://doi.org/10.1016/j.asw.2016.03.001>
- Damayanti, F. Y., Setyaningsih, E., & Sumardi, S. (2024). Teachers' language assessment literacy: Indonesian secondary EFL teachers' knowledge, perspective, and classroom practice. *Tarbiyah: Jurnal Ilmiah Kependidikan*, 13(2), 171–185. <https://doi.org/10.18592/tarbiyah.v13i2.13602>
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415. <https://doi.org/10.1002/j.1545-7249.2009.tb00242.x>
- Derakhshan, A., & Ghiasvand, F. (2022). Demystifying Iranian EFL teachers' perceptions and practices of learning-oriented assessment (LOA): Challenges and prospects in focus. *Language Testing in Asia*, 12(1), 55. <https://doi.org/10.1186/s40468-022-00204-2>
- Derakhshan, A., Shakki, F., & Sarani, M. A. (2020). The effect of dynamic and non-dynamic assessment on the comprehension of Iranian intermediate EFL learners' speech acts of apology and request. *Language Testing in Asia*, 12(1), 55.
- Direktorat Jenderal Guru dan Tenaga Kependidikan. (2023). *Peraturan Direktur Jenderal Guru dan Tenaga Kependidikan Nomor 4560/B/HK.03.01/2023 tentang Petunjuk Teknis Pelaksanaan Program Pendidikan Profesi Guru Prajabatan*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Direktorat Jenderal Guru, Tenaga Kependidikan, dan Pendidikan Guru. (2024). *Perdirjen GTK Nomor 441 Tahun 2024 tentang Pedoman Penyelenggaraan Pendidikan Profesi Guru*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. <https://ppg.kemendikdasmen.go.id/storage/2025/Perdirjen%20GTK%20Nomor%20441%20Tahun%202024%20tentang%20Pedoman%20Penyelenggaraan%20PPG.pdf>
- Duff, P. A. (2018). *Case study research in applied linguistics*. Routledge.
- Estaji, M., & Ghiasvand, F. (2021). Assessment perceptions and practices in academic domain: The design and validation of an assessment identity questionnaire (TAIQ) for EFL teachers (Pt. 103-131). *International Journal of Language Testing*, 11(1).
- Farmasari, S., Wardana, L. A., Baharudin, B., Herayana, D., & Suryaningsih, H. (2023). Pre-service EFL teachers' language assessment literacy satisfaction and assessment preparedness. *International Journal of Language Education*, 7(4). <https://doi.org/10.26858/ijole.v7i4.47445>
- Fitriyah, I., Masitoh, F., & Widiati, U. (2022). Classroom-based language assessment literacy and professional development need between novice and experienced EFL teachers. *Indonesian Journal of Applied Linguistics*, 12(1), 124–134. <https://doi.org/10.17509/ijal.v12i1.46539>

- Fulcher, G. (2003). *Testing second language speaking*. Routledge. <https://doi.org/10.4324/9781315837376>
- 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. (2015). Assessing second language speaking. *Language Teaching*, 48(2), 198–216. <https://doi.org/10.1017/S0261444814000391>
- Fulcher, G. (2021). Language assessment literacy in a learning-oriented assessment framework. In *Learning-oriented language assessment: Putting theory into practice* (pp. 34–48). Routledge. <https://doi.org/10.4324/9781003014102>
- Gan, Z., Leung, C., He, J., & Nang, H. (2019). Classroom assessment practices and learning motivation: A case study of Chinese EFL students. *TESOL Quarterly*, 53(2), 514–529. <https://doi.org/10.1002/tesq.476>
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>
- Hakim, L. N., Widiati, U., Purwaningsih, S., Wulyani, A. N., & Mulyanti, W. (2024). Revisiting the content and instruction of TEFL methodology course: A needs analysis. *Pegem Journal of Education and Instruction*, 14(1), 167–177. <https://doi.org/https://doi.org/10.47750/pegegog.14.01.19>
- Inayah, N., Komariah, E., & Nasir, A. (2019). The practice of authentic assessment in an EFL speaking classroom. *Studies in English Language and Education*, 6(1), 152–162. <https://doi.org/10.24815/siele.v6i1.13069>
- 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>
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. (2022). *Permendikbudriset No. 21 Tahun 2022 tentang Standar Penilaian Pendidikan pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah—JDIH BPK*. Database Peraturan | JDIH BPK. <https://peraturan.bpk.go.id/Details/224425/permendikbudriset-no-21-tahun-2022>
- Koh, K., Burke, L. E. C.-A., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264–288. <https://doi.org/10.1177/1362168816684366>
- Kremmel, B., & Harding, L. (2020). 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>
- 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>
- Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, 61, 112–122. <https://doi.org/10.1016/j.stueduc.2019.03.003>
- 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>
- Li, J., & Gu, P. Y. (2023). Developing classroom-based formative assessment literacy: An EFL teacher's journey. *Chinese Journal of Applied Linguistics*, 46(2), 198–218. <https://doi.org/10.1515/CJAL-2023-0204>
- Luoma, S. (2004). *Assessing speaking* (6. print). Cambridge University Press.

- Luthfiyyah, R., Basyari, I. W., & Dwiniasih, D. (2020a). EFL secondary teachers' assessment literacy: Assessment conceptions and practices. *Journal on English as a Foreign Language*, 10(2), 402–421. <https://doi.org/10.23971/jevl.v10i2.2101>
- Luthfiyyah, R., Basyari, I. W., & Dwiniasih, D. (2020b). EFL secondary teachers' assessment literacy: Assessment conceptions and practices. *Journal on English as a Foreign Language*, 10(2), 402–421. <https://doi.org/10.23971/jevl.v10i2.2101>
- McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction* (5th ed). Pearson.
- Möller, A., Comer, K., Rawlins, P., & Emerson, L. (2025). Using a scenario-based method to explore university students' views about writing assistance. *New Zealand Journal of Educational Studies*, 60(1), 225–245. <https://doi.org/10.1007/s40841-025-00384-4>
- Norhasim, N. A., & Mohamad, M. (2020). Teachers' challenges on the implementation of classroom-based assessment for oral proficiency: A literature review. *Creative Education*, 11(10), 1922–1929. <https://doi.org/10.4236/ce.2020.1110140>
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–12. <https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>
- Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly*, 43(3), 471–491. <https://doi.org/10.1002/j.1545-7249.2009.tb00245.x>
- PPG bagi Calon Guru—Direktorat Pendidikan Profesi Guru (PPG). (n.d.). Retrieved September 25, 2026, from <https://ppg.kemendikdasmen.go.id/ppg-calon-guru>
- Prastikawati, E. F., Mujiyanto, Y., Saleh, M., & Fitriati, S. W. (2024). Teaching practicum: Does it impact to pre-service EFL teachers' language assessment literacy? *World Journal of English Language*, 14(3), 242. <https://doi.org/10.5430/wjel.v14n3p242>
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190–208. <https://doi.org/10.1111/modl.12308>
- Pusat Asesmen Pendidikan. (2021, April 19). *Penjelasan Mendikbud Terkait 3 Aspek Asesmen Nasional Pengganti UN 2021*. https://pusmendik.kemendikdasmen.go.id/konten/penjelasan-mendikbud-terkait-3-aspek-asesmen-nasional-pengganti-un-2021?utm_source=chatgpt.com
- Qotboddin, J. M., Hossein, K., & Khalil, M. (2020). Assessment literacy of Iranian EFL teachers: A review of recent studies. *The Journal of AsiaTEFL*, 17(2), 689–698. <https://doi.org/10.18823/asiatefl.2020.17.2.27.689>
- Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 183–200. <https://doi.org/10.1177/0002716219843818>
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534–539.
- Stravakou, P., & Lozgka, E. (2018). Vignettes in qualitative educational research: Investigating Greek school principals' values. *The Qualitative Report*, 23(5), 1188–1207. <https://doi.org/10.46743/2160-3715/2018.3358>
- 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>
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 1. <https://doi.org/10.1186/s40468-019-0077-8>
- Sumardi, S., & Guci, R. I. (2023). HOTS-based language assessment literacy: Challenges and prospects in English language teaching. *Indonesian Journal of Applied Linguistics*, 12(3), 831–840. <https://doi.org/10.17509/ijal.v12i3.44261>
- Tajeddin, Z., Alemi, M., & Yasaei, H. (2018). Classroom assessment literacy for speaking: Exploring novice and experienced English language teachers' knowledge and

- practice. *Iranian Journal of Language Teaching Research*, 6(3), 57–77. <https://files.eric.ed.gov/fulltext/EJ1192613.pdf>
- 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., & 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>
- Turner, C. E., & Purpura, J. E. (2016). 16 Learning-oriented assessment second and foreign classrooms. In *Handbook of Second Language Assessment* (In D. Tsagari&J. Banerjee (Eds.), pp. 255–274). e Gruyter Mouton. <https://doi.org/10.1515/9781614513827-018>
- 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>
- Vogt, K., Tsagari, D., Csépes, I., Green, A., & Sifakis, N. (2020). Linking learners’ perspectives on language assessment practices to teachers’ assessment literacy enhancement (TALE): Insights from four European countries. *Language Assessment Quarterly*, 17(4), 410–433. <https://doi.org/10.1080/15434303.2020.1776714>
- Wang, X. (2017). A Chinese EFL teacher’s classroom assessment practices. *Language Assessment Quarterly*, 14(4), 312–327. <https://doi.org/10.1080/15434303.2017.1393819>
- 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., & Fan, J. (2021). “Am I qualified to be a language tester?”: Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246. <https://doi.org/10.1177/0265532220929924>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (Sixth edition). SAGE.
- Zulaiha, S., & Mulyono, H. (2020). Exploring junior high school EFL teachers’ training needs of assessment literacy. *Cogent Education*, 7(1), 1772943. <https://doi.org/10.1080/2331186X.2020.1772943>
- 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>