



---

**English Language Assessment Practices in Batam Secondary Schools: Teachers' Strategies, Constraints, and Pedagogical Negotiations**Oey anton<sup>1</sup>, Mariati<sup>2</sup>, Alexandria<sup>3</sup><sup>1,2,3</sup>Universitas Universal, Batam, IndonesiaCorresponding author email: [oyey.anton@uvers.ac.id](mailto:oyey.anton@uvers.ac.id)

---

**Info Article**

Received: 8 May 2026

Revised: 15 May 2026

Accepted: 16 May 2026

OnlineVersion: 18 May 2026

---

**Abstract**

This study examines English language assessment practices in Batam secondary schools by focusing on how teachers balance pedagogical goals, institutional requirements, and classroom constraints. The study employed a qualitative descriptive design using interview records from English teachers representing several school contexts in Batam, including junior high school, senior high school, vocational school, and independent school settings. The data were analyzed thematically across recurrent domains such as oral and written assessment, formative and diagnostic assessment, feedback practices, project and quiz use, summative examination structure, grading systems, assessment media, and classroom challenges. The findings indicate that teachers generally value balanced, authentic, and student-centered assessment, yet written assessment remains more dominant because it is more practical in large classes and easier to align with formal school reporting. Oral assessment is still considered indispensable because it reveals confidence, fluency, pronunciation, and spontaneous language use that written work often cannot capture. The study also shows that formative assessment is commonly implemented through quizzes, games, classroom interaction, worksheets, projects, and short performance tasks, while summative assessment tends to prioritize reading, grammar, vocabulary, and writing. Digital tools such as Google Forms, Quizizz, Wordwall, ZEPP, Exambro, and school-based systems are used selectively, but teachers continue to rely on paper-based records and direct classroom observation to ensure fairness and authenticity. The paper argues that assessment in Batam secondary schools reflects an ongoing negotiation between authenticity and manageability, and that future improvement should strengthen performance-based assessment, clearer rubrics, and practical institutional support for teachers.

Keywords: English language assessment; formative assessment; summative assessment; Batam; EFL

*This is open access article under the [CC-BY](https://creativecommons.org/licenses/by/4.0/) licence*



---

**INTRODUCTION**

Assessment is one of the most influential components of English language teaching because it shapes what teachers teach, how students learn, and how achievement is interpreted in school contexts. In English as a foreign language setting, assessment is no longer viewed only as a tool for giving scores at the end of instruction, but also as an integral process for monitoring progress, diagnosing needs, and

guiding improvement throughout learning. Recent scholarship has highlighted that formative assessment can support classroom interaction, feedback, learner engagement, and adaptive teaching, especially when it is embedded into everyday instructional practice rather than restricted to formal testing alone (Zhang et al., 2024).

At the same time, assessment in real classrooms is rarely designed under ideal conditions. English teachers often need to reconcile multiple demands at once: curriculum expectations, school grading systems, time limits, class size, administrative duties, student motivation, digital technology, and concerns about fairness. Previous studies have shown that although innovative forms such as technology-assisted assessment, peer assessment, and performance-based tasks can enrich learning, teachers still face practical and cultural barriers in implementing them consistently in EFL contexts (Zhang et al., 2024). Research on assessment literacy also suggests that teachers frequently need stronger support in translating assessment principles into workable classroom practice, especially when trying to balance validity, efficiency, and student needs (Hussin et al., 2022; Putri et al., 2024).

Within Indonesia, English assessment has also been shaped by curriculum reform, school policy, and increasing pressure to make learning more authentic and relevant. However, local implementation remains highly context-dependent. School type, student profile, teacher experience, available technology, and institutional policy may strongly influence how assessment is interpreted and enacted. This means that assessment practice in one region cannot simply be assumed to look the same in another. Batam offers a particularly useful context for examining this issue because its schools include diverse institutional forms, such as vocational schools, independent schools, junior high schools, and senior high schools, each of which may prioritize English differently.

The interview corpus used in this study shows that teachers in Batam employ a wide range of assessment methods, including written tests, oral tasks, projects, interviews, quizzes, diagnostic checks, direct observation, games, worksheets, and digital platforms. However, these methods are not applied with equal intensity or for identical purposes. Some teachers rely more strongly on written assessment because it is easier to score and document. Others actively attempt to preserve oral and performance-based assessment because they believe these forms better reveal students' real communicative ability. Teachers also vary in how they design projects, provide feedback, manage large classes, and weigh formative and summative components in final grades.

This study therefore seeks to answer the following questions. First, how do English teachers in Batam balance oral and written assessment in secondary schools? Second, how do they implement formative, diagnostic, and feedback-based assessment for vocabulary, grammar, and broader language development? Third, how are summative examinations, grading systems, and assessment tools organized in school practice? Fourth, what challenges and ideal assessment visions emerge from teachers' experiences across school contexts? By addressing these questions, the paper contributes a context-rich account of how classroom assessment is negotiated in Batam and how that negotiation reflects broader tensions in EFL assessment between authenticity, accountability, and practicality.

A locally relevant study also strengthens the rationale for examining assessment in Batam. Anton and Yodi (2021) used the EnglishScore Android application to map the CEFR-based proficiency of 155 Universitas Universal learners in Batam and found that most learners clustered at A2 and B1 levels, with reading and listening still needing improvement. Although that study focused on proficiency prediction rather than classroom assessment practices, it demonstrates that digital assessment tools and CEFR-based profiling are already part of the wider Batam educational landscape. This makes the present study more contextually grounded because the teachers interviewed in Batam likewise reported increasing engagement with digital tools, diagnostic assessment, and practical ways of identifying learner ability.

This study offers novelty in two ways. First, it provides a cross-context qualitative account of English assessment practices in Batam by comparing teachers from independent, junior high, senior high, and vocational school settings rather than focusing on a single institution or a single assessment type. Second, it shows that assessment in EFL classrooms is not merely a technical matter of choosing test formats, but a situated pedagogical negotiation shaped by authenticity, manageability, school policy, class size, technology access, and teacher judgment. Theoretically, the study contributes to discussions of

authentic assessment and teacher assessment literacy by showing that teachers' classroom decisions are influenced not only by assessment principles, but also by local institutional ecology and practical constraints.

## RESEARCH METHODS

### *Research Design*

This study employed a qualitative descriptive design. The primary data consisted of interview transcripts collected from nine English teachers working in secondary school contexts in Batam through purposive sampling. The interview materials covered junior high school, senior high school, vocational high school, and independent school settings, which enabled cross-context comparison of assessment practices. Because some schools were represented by more than one interview transcript, the analysis treated the dataset as a thematic corpus rather than as a simple one-file-one-case dataset.

### *Research Target/Subject*

The interview questions addressed recurring domains related to English classroom assessment, namely: (1) oral and written assessment balance, (2) vocabulary and grammar assessment, (3) formative, project-based, and quiz-based assessment, (4) diagnostic assessment, (5) feedback practices, (6) summative exam structure, (7) grading systems and school policy, (8) media and digital tools, (9) classroom and administrative challenges, and (10) teachers' views of ideal assessment systems. These domains provided a stable analytical frame across the interview dataset.

This study involved nine English teachers from different secondary school contexts in Batam. The participants were selected through purposive sampling because they met the main criteria of this study: they taught English in Batam secondary schools and had direct experience with classroom assessment practices. To maintain confidentiality, all participants are presented using anonymized codes, namely T1 to T9.

Table 1. The characteristics of the participants are presented

<b>Code</b>	<b>School context</b>	<b>School level taught</b>	<b>Teaching experience</b>
T1	Independent school in Batam	Secondary-school	5 years
T2	Putra Jaya School Batam	Vocational high school	8 years
T3	SMA Bodhidharma Batam	Senior high school	15 years
T4	SMK Maitreyawira Batam	Vocational high school	About 1 year in formal school teaching; around 4 to 5 years overall
T5	SMA Maitreyawira Batam	Senior high school	15 years in junior and senior high school
T6	Permata Harapan Batam	Junior and senior high school	Around 8 years
T7	SMA Kristen Tabqha	Senior high school	Around 3 years
T8	SMK Kartini Batam	Vocational high school	4 years
T9	Putra Batam School	Junior high school	5 years

### *Research Procedure*

The data were analyzed using thematic analysis. First, all interview data were reviewed repeatedly to identify recurring ideas and school-specific practices. Second, statements were grouped into broader

categories such as authenticity, practicality, formative monitoring, grading fairness, technology use, and classroom constraints. Third, cross-case comparison was conducted to identify both common patterns and distinctive practices. Finally, the emergent themes were interpreted in relation to current scholarship on formative assessment, authentic assessment, and teacher assessment literacy in EFL contexts. The purpose of this approach was not statistical generalization, but analytic description of how teachers negotiate assessment practice under real school conditions.

To ensure trustworthiness, this study applied member checking and peer debriefing. After the preliminary themes were identified, selected participants were asked to confirm whether the interpretations accurately reflected their views and classroom assessment practices. Peer debriefing was also conducted with academic peers to review the coding process, question possible researcher bias, and refine the thematic categories. In addition, credibility was strengthened through cross-case comparison across teachers from different school contexts, while dependability was supported by using a consistent analytical framework across the interview transcripts. These procedures helped make the interpretation more credible, transparent, and contextually grounded.

### ***Instruments, and Data Collection Techniques***

To improve trustworthiness, themes were not drawn from a single transcript only, but were retained when supported by repeated evidence across different school settings. At the same time, unusual or school-specific practices, such as leaderboard systems, essay-only exam policies, gallery walk speaking assessment, workplace-oriented vocational tasks, and smartphone-based exam systems, were also preserved because they provided contextually important insight into how English assessment is adapted in Batam schools.

### ***Data analysis technique***

The data were interpreted through thematic analysis by identifying recurring patterns across teacher interviews and relating these patterns to the study objectives and relevant literature.

## **RESULTS AND DISCUSSION**

The findings are presented in four major themes: the dominance of written assessment alongside the continuing importance of speaking, formative assessment through quizzes and projects, assessment of vocabulary and grammar through contextualized tasks, and the tension between authenticity and practicality in real classroom conditions. Rather than treating each school as a separate case, the discussion emphasizes recurring patterns across contexts while using selected examples only when they clarify distinctive practices.

### ***Dominance of Written Assessment and the Continuing Importance of Speaking***

The Batam interview corpus consistently shows that teachers value both oral and written assessment, but written assessment remains more dominant in routine practice. Across independent, junior high, senior high, and vocational school settings, teachers explained that written tasks are easier to administer, easier to document, and more manageable under limited instructional time. This tendency appeared in independent-school data where one teacher estimated the balance at around 70 percent writing and 30 percent oral, in SMP Maitreyawira where written tests were preferred for practical reasons, and in several SMK Maitreyawira responses where writing occurred more frequently because it naturally accompanied reading, grammar, and essay tasks.

Even so, teachers did not dismiss speaking as secondary or optional. On the contrary, many respondents described speaking as essential for assessing confidence, fluency, pronunciation, spontaneity, and real communicative ability. Teachers in SMK Maitreyawira emphasized that speaking could reveal aspects that writing could not, especially because written work might be supported by dictionaries, peers, or AI-based tools. In one transcript, oral assessment was described as more authentic precisely because students could not easily fake their competence when speaking in real time. These findings align with

broader discussions of authentic assessment, which emphasize performance evidence as a meaningful way to capture actual language use rather than only decontextualized textual accuracy (Murtiana, 2024; Authentic Assessment in EFL Classrooms, 2024).

However, the prominence of written assessment does not simply indicate teacher preference for traditional testing. Rather, it reflects structural conditions. Speaking assessment often requires one-by-one observation, more class time, and stronger classroom control. Teachers therefore attempt to preserve oral tasks while limiting their frequency or embedding them into selected chapters, projects, interviews, role-plays, presentations, storytelling, or performance sessions. This pattern supports the idea that assessment practice is shaped not only by beliefs about language learning but also by feasibility under real school constraints (Zhang et al., 2024).

### ***Formative Assessment Through Quizzes, Games, Projects, and Daily Interaction***

Teacher accounts strongly indicate that formative assessment is widely practiced in Batam, although its forms vary across schools. Teachers described using quizzes, worksheets, class discussions, short assignments, oral questioning, project tasks, participation monitoring, and observation of daily classroom activity. In some schools, quizzes were conducted at the end of a unit; in others, they were used every few weeks or even embedded into each chapter. Several vocational and junior high school teachers explained that formative assessment carried substantial weight in final grading, often 40 percent, because daily performance was considered a more stable indicator than one high-stakes exam alone.

Game-based formative assessment emerged as a particularly visible pattern, especially in SMK Maitreyawira and SMP Maitreyawira. Teachers reported using spelling activities, guessing games, crossword puzzles, taboo-like tasks, Wordwall, ZEPP, and competitive quiz formats to reduce pressure and increase engagement. These tools were not used merely for entertainment. Teachers described them as ways of checking vocabulary, grammar, responsiveness, and participation while maintaining student motivation. This supports recent research showing that formative assessment in EFL contexts can be enhanced by interactive and technology-assisted methods, provided they remain aligned with learning goals rather than functioning as isolated activities (Zhang et al., 2024; AI-Mediated Digital Formative Assessment, 2026).

Projects were also important, but teachers used them more selectively than quizzes. In some schools, projects were assigned once or twice per semester, while in others project work appeared in each section or chapter. Teachers justified project use when tasks matched the material well, such as recounts, posters, presentations, interviews, drama, or vocationally relevant products. Yet project use was often limited by time, academic calendar pressure, and concern that group work could hide individual weakness. This ambivalence is significant. It shows that teachers value projects for deeper application and engagement, but still worry about fairness and evidence quality. Similar tensions are noted in studies of authentic assessment, where realistic tasks are pedagogically valuable but harder to monitor, score, and standardize consistently (Murtiana, 2024).

### ***Assessing Vocabulary, Grammar, and Prior Knowledge***

The data suggest that Batam teachers assess vocabulary and grammar in relatively contextualized ways, even when they still rely on written formats. Teachers mentioned fill-in-the-blank exercises, paragraph writing, sentence analysis, vocabulary diaries, highlighted words in reading texts, synonym and antonym tasks, editing exercises, puzzles, direct questioning, and contextual grammar items. Bodhidharma data show vocabulary being integrated into reading passages from the English in Mind textbook, while independent-school and vocational-school teachers described a combination of word-form tasks, sentence-based practice, and short contextual writing. This indicates that vocabulary and grammar are generally not treated as purely isolated memorization targets, even when formal testing remains prominent.

Diagnostic assessment was also visible across multiple schools. Teachers used pre-tests, oral questioning, short quizzes, four-skill introductions at the beginning of the semester, and chapter-opening checks related to prerequisite knowledge. The purpose was not only to label students but to decide where instruction should begin. For example, one teacher explained that passive voice could not be taught meaningfully unless students already understood basic tenses and forms of to be. Others used diagnostic results for grouping students, assigning peer support, or differentiating follow-up tasks. Such practices are

consistent with the idea that formative and diagnostic assessment should inform subsequent teaching rather than function as isolated measurement events (Zhang et al., 2024).

At the same time, teachers openly acknowledged that vocabulary and grammar assessment remains challenging. Some learners may appear to understand orally but struggle in writing; others can complete written work but still cannot apply language spontaneously. This reinforces the need for multiple forms of evidence. Research on teacher assessment literacy similarly argues that teachers need both conceptual and practical competence to select assessment modes that genuinely represent learner ability rather than relying on one format out of habit or convenience (Hussin et al., 2022; Putri et al., 2024).

### ***Feedback, Remedial Practice, and Support for Student Progress***

Feedback was one of the most consistently recurring themes in the Batam interview records. Teachers reported using both written and oral feedback, depending on the task and the immediacy of the problem observed. Written feedback was typically given on notebooks, worksheets, essays, or structured tasks, whereas oral feedback was more common after speaking performances, interviews, presentations, or class interaction. Some teachers described direct correction, while others preferred delayed discussion after a student had finished speaking to avoid embarrassment. This variation suggests that feedback was not treated as a single fixed technique but as a flexible pedagogical response tied to task type and student condition.

Several interview records also show that teachers understood feedback as inseparable from student encouragement. Bodhidharma data emphasized remedial progression, where a student could still be seen as progressing even when not yet meeting the passing standard because scores improved over repeated formative checks. SMK Maitreyawira teachers described reflection sessions, revision after feedback, and motivational support for students who were passive, discouraged, or afraid of making mistakes. Putra Jaya School data similarly stressed guidance and personal approach, particularly for weaker learners. These findings resonate with literature indicating that formative feedback functions most effectively when it combines cognitive guidance with motivational and socio-emotional support (Zhang et al., 2024; Exploring EFL Teachers' Formative Assessment Literacy, 2024).

The importance of feedback in the Batam data also illustrates a broader pedagogical belief: assessment should contribute to improvement rather than merely sort students by score. Some teachers explicitly argued that assessment is also a form of teacher reflection because weak results may indicate not only student difficulty but also a need to adjust explanation, pacing, or task design. This perspective is notable because it moves assessment beyond accountability and closer to responsive pedagogy, a point repeatedly emphasized in contemporary formative assessment research (Zhang et al., 2024).

SMA Maitreyawira data additionally show that feedback practices can be selective and situational rather than routine. The teacher acknowledged that feedback was sometimes skipped and that spoken feedback was used less often because reading students' written reflections was faster and allowed input from more learners. At the same time, she used project journals and occasional post-project questioning to understand student experience and adjust later tasks. This confirms that feedback in Batam classrooms may serve both evaluative and reflective purposes, but its consistency still depends heavily on planning and workload.

### ***Summative Assessment, Grading Systems, and the Logic of Practicality***

Summative assessment in Batam secondary schools generally remains more conventional than formative assessment. Teachers described midterm and final examinations as focusing mainly on reading, grammar, vocabulary, writing, and sometimes listening, while speaking was frequently separated from formal semester exams because of time and logistical constraints. In vocational and senior high school settings, listening could be included through paper-based or technology-supported formats, but one-by-one speaking remained difficult to integrate into high-stakes schedules. This pattern confirms a clear division in practice: speaking is often acknowledged as pedagogically important, yet still marginalized in formal summative structures when schools prioritize manageability.

Exam design was commonly described as aligned with previously taught material, which suggests a strong concern for fairness and curricular fit. Teachers referred to essay items, multiple-choice and short-

answer combinations, fill-in-the-blank tasks, editing items, text analysis, and even IELTS-like reading formats such as matching headings. Putra Jaya School stood out for its essay-only policy, whereas Bodhidharma described a fixed 10 multiple-choice plus 5 completion-item structure. Independent-school data indicated large written exams with editing, vocabulary, grammar, reading, and summary writing sections. Across these variations, teachers consistently emphasized that tests should reflect material already introduced in class.

Grading systems also reveal a negotiation between school policy and teacher judgment. Some schools reported a 40-30-30 distribution for daily work, midterm, and final exams; others emphasized 20-30-50 or heavier weighting for final summative components. Teachers also mentioned quizzes, daily tasks, projects, attitude, participation, and practical tasks as part of final reporting. Importantly, several respondents questioned whether some scores truly reflected individual ability, especially in group projects or AI-assisted work. This concern with fairness and validity is central to assessment literacy, since grading becomes problematic when numerical results no longer accurately represent actual competence (Hussin et al., 2022; Language Assessment Literacy in Online EFL Reading Class, 2025).

Additional interview evidence adds particularly strong support to the discussion of summative policy and grading logic. The teacher stated that midterm and final examinations were compulsory paper-based cognitive tests determined by the school system, while daily assessment outside those major exams could still be managed more flexibly through spoken tasks, projects, or other teacher-designed methods. She also indicated that the school controlled a major portion of the final score through midterm and final exams, while the remaining part could be shaped by the teacher's own assessment choices. This provides clearer support for the argument that teachers work within partially fixed institutional assessment structures rather than designing the whole system independently.

### ***Technology Use, Fairness, and the Problem of Authentic Evidence***

This pattern is also consistent with earlier Batam-based evidence showing that digital testing has been used to classify learner proficiency through CEFR-aligned tools such as EnglishScore, which suggests that technology-mediated assessment is not new in the local context but is evolving from proficiency mapping toward broader classroom use (Anton & Yodi, 2021).

The Batam interview records portray a mixed assessment ecology in which digital and non-digital tools coexist rather than replace one another. Teachers mentioned Google Forms, Quizizz, Wordwall, ZEPP, Exambro, YouTube, school LMS platforms, online spreadsheets, WiFi-supported projector lessons, and traditional paper-based tests or notebook records. Digital tools were generally valued for speed, engagement, score calculation, or interactive quiz design. In some schools, they were especially useful for listening sections, competitive quizzes, or automatic scoring of objective items. In others, they were used more cautiously because essay responses, speaking evaluation, and fairness concerns still required teacher judgment.

Teachers did not present technology as uniformly positive. Bodhidharma data pointed out that Google Forms was less suitable for essay-heavy tasks. SMK Maitreyawira interview records described internet instability, non-premium limitations in Wordwall, and Exambro login delays that disrupted listening sections. In the independent-school data, the school LMS was still under development, so teachers relied more on Quizizz. Meanwhile, some respondents argued that digital platforms increase focus and engagement, while others worried that computer-based testing encourages shortcut thinking or easier access to AI-generated responses. This ambivalence closely matches recent discussions on digital formative assessment, where technological efficiency can improve participation but also creates new validity and ethics questions in EFL writing and classroom assessment (AI-Mediated Digital Formative Assessment, 2026; Language Assessment Literacy in Online EFL Reading Class, 2025).

The issue of authenticity was especially visible in teacher responses to AI use and take-home writing. Several respondents noted that AI-assisted or heavily supported writing may look polished but does not necessarily represent the student's own competence. For this reason, some teachers preferred in-class writing, speaking tasks, open-ended questions, direct interviews, or process observation during project work. In other words, the Batam data show that authenticity is not an abstract concept for teachers. It is a

practical concern linked to cheating, over-assistance, and the difficulty of knowing whose language performance is actually being assessed. This finding echoes research on authentic assessment, which argues that real evidence of learning must involve meaningful performance while remaining sufficiently observable and attributable to the learner (Murtiana, 2024; *Teacher's Beliefs and Practices of Authentic Assessment in EFL*, 2024).

### ***Classroom Constraints and Teachers' Ideal Assessment Systems***

Although practices differed across schools, teachers repeatedly described similar constraints. The most common were time limitations, grading burden, large classes, speaking-assessment logistics, student passivity, and administrative work. Speaking was often the hardest skill to assess because it required direct attention to individual learners while the rest of the class also needed to stay engaged. Teachers therefore developed local strategies such as gallery walk, peer assessment, rotating performance spots, rubrics, spinning-wheel presentation turns, bonus-point quizzes, and classroom movement activities. These strategies show that teachers are not passively constrained by conditions; they actively improvise solutions to preserve assessment quality under pressure.

Administrative burden was reported differently across schools. Some teachers felt heavily burdened by lesson planning, score compilation, and physical documentation, especially in vocational school contexts where printed records were still required. Others, such as at Bodhidharma, described online spreadsheet systems as reducing workload because scores could be entered directly and analyzed automatically. These differences indicate that institutional infrastructure significantly affects how burdensome assessment becomes. Assessment quality is therefore not only a matter of teacher competence but also of organizational support and system design.

When asked about ideal assessment systems, teachers generally did not advocate removing written exams altogether. Instead, they wanted a more balanced model that retained written tests while increasing authentic, performance-based, and daily-life-relevant assessment. Vocational-school teachers wanted stronger alignment with workplace communication, such as CV writing, job interviews, application letters, and practical English use. Other teachers emphasized daily performance, participation, role-play, storytelling, or interview-based speaking. Taken together, these responses suggest that Batam teachers envision assessment as most meaningful when it combines accountability with relevance, and when it measures not only what students can reproduce on paper but also what they can actually do with English in real contexts.

## **CONCLUSION**

This study demonstrates that English language assessment in Batam secondary schools is best understood as a process of pedagogical negotiation rather than a fixed technical system. Teachers across school contexts value balance, authenticity, and student development, yet their daily choices are strongly shaped by time, class size, grading structures, institutional rules, and technological limitations. As a result, written assessment remains dominant not because teachers necessarily consider it superior in principle, but because it is more manageable under prevailing school conditions.

At the same time, the findings show that teachers actively try to preserve more meaningful forms of assessment. They use oral tasks, projects, games, diagnostics, quizzes, interviews, reflection, remedial feedback, and digital tools to broaden the evidence they collect about students' English ability. Their practices reveal both creativity and tension: creativity in adapting assessment to local realities, and tension in trying to measure authentic performance without losing fairness, efficiency, or curricular alignment.

The implication is that improving assessment in Batam secondary schools should not mean replacing written tests with idealized alternatives. Instead, schools and policymakers should support realistic improvement through workable rubrics, better digital infrastructure, stronger teacher assessment literacy, manageable speaking-assessment scheduling, and assessment models that recognize both process

and performance. Such support would help teachers move closer to the kind of balanced, student-centered, and contextually relevant English assessment that they themselves already envision.

### ACKNOWLEDGMENTS

The authors would like to thank all teacher participants and schools that contributed interview data to this study. Appreciation is also extended to colleagues and institutions that supported the preparation of this manuscript.

### REFERENCES

- AI-Mediated Digital Formative Assessment. (2026). AI-mediated digital formative assessment in EFL writing: Validity, literacy, and ethics. *Journal of English Language Education*, 11(1).
- Anton, O., & Yodi. (2021). Predicting English language learners' proficiency level using EnglishScore Android application. *Journal of Digital Ecosystem for Natural Sustainability*, 1(2), 106–110. <https://journal.uvers2.ac.id/index.php/jodens>
- Authentic Assessment in EFL Classrooms. (2024). Authentic assessment in EFL classrooms. *Jurnal Joepallt*, 12(2).
- Hussin, H., Yusoff, N. M., & Nimehchisalem, V. (2022). ESL teachers' assessment literacy in classroom: A review of past studies. *Journal of Language and Linguistic Studies*, 18(1), 112–127.
- Language Assessment Literacy in Online EFL Reading Class. (2025). Language assessment literacy in online EFL reading class. *Computer-Assisted Language Learning Electronic Journal*, 26(2).
- Murtiana, R. (2024). Authentic assessment in examining EFL students' writing. *Eternal (English Teaching Journal)*, 15(1).
- Putri, D., Rahman, A., & colleagues. (2024). Teachers' language assessment literacy: Indonesian secondary school EFL contexts. *Journal of Teaching and Learning in Indonesia*, 6(2).
- Teacher's Beliefs and Practices of Authentic Assessment in EFL. (2024). Teacher's beliefs and practices of authentic assessment in EFL speaking classrooms. *Journal of Innovation in Language Education*, 4(2).
- Zhang, H., Ge, S., & Saad, M. R. B. M. (2024). Formative assessment in K-12 English as a foreign language education: A systematic review. *Heliyon*, 10(10), e31367. <https://doi.org/10.1016/j.heliyon.2024.e31367>

