



Integrating Assessment Tools and Practices in EFL Classrooms: A Multidimensional Approach Aligned with CEFR Standards

Asep Ahmad ZA^{1*}, Ratmo¹

¹Universitas Pamulang

*dosen00480@unpam.ac.id

ABSTRACT

This study critically examines the evolving role of assessment and evaluation in Teaching English as a Foreign Language (EFL), emphasizing a paradigm shift from summative judgment to a continuous, formative process integral to instruction. Through qualitative content analysis of theoretical literature and practical frameworks, the research synthesizes multidimensional assessment practices that align with the Common European Framework of Reference for Languages (CEFR). It explores diverse assessment tools—including written exams, oral tests, multiple-choice questions, performance tasks, checklists, and analytic rubrics—and evaluates their efficacy in measuring integrated language skills (listening, speaking, reading, writing). Findings reveal that imbalances in skill development often stem from isolated assessment methods, leading to communicative inadequacies (e.g., comprehension without production). The study advocates for contextually authentic, CEFR-aligned strategies that bridge receptive and productive competencies, such as combining reading with written responses or listening with oral tasks. Additionally, it underscores the distinction between assessment (systematic observation) and evaluation (criterion-referenced judgment), highlighting the need for balanced tool selection to enhance reliability and reduce anxiety. Practical implications include leveraging checklists for skill-specific benchmarks, analytic rubrics for performance clarity, and iterative feedback for writing development. The research concludes that holistic, integrated assessment designs are essential for fostering balanced language proficiency and achieving meaningful educational outcomes in global EFL contexts.

This is an open access article under [CC-BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/) license.



ARTICLE INFO

Keywords:

assessment
CEFR;
Formative process;
Summative judgment;
Teaching English

How to Cite in APA Style:

Ratmo, & ZA, A. A. . (2025). Integrating Assessment Tools and Practices in EFL Classrooms: A Multidimensional Approach Aligned with CEFR Standards. *IJLHE: International Journal of Language, Humanities, and Education*, 8(1), 247–264.
<https://doi.org/10.52217/ijlhe.v8i1.1823>

INTRODUCTION

Education, in its most general sense, is the process of guiding individuals toward predetermined objectives. By the end of this process, it is expected that individuals will have undergone measurable transformation, having achieved the

intended goals. To determine whether, and to what extent, these objectives have been met, assessment and evaluation practices are employed. According to O'Dwyer & de Boer, (2015), such practices play a crucial role not only in identifying the current status of learners but also in steering the educational process toward the attainment of its established aims.

Historically, assessment and evaluation have always held a position of significance within the educational enterprise. However, the functions attributed to them have evolved over time, expanding in scope and gaining increased importance in contemporary pedagogical contexts. Initially, assessment was perceived as a summative tool—an activity conducted at the conclusion of the learning process to determine student achievement. The primary aim of such evaluations was to categorize learners as either “successful” or “unsuccessful,” typically based on a single measurement taken after instruction had ended (Yüce & Mirici, 2023)

In contrast, contemporary educational paradigms no longer confine assessment to the narrow task of measuring a learner's level of attainment. Today, assessment and evaluation are understood to serve multiple functions: they facilitate a comprehensive understanding of students, inform the planning and execution of instructional activities, and help gauge the effectiveness of every component within the learning process. As such, assessment is now viewed not merely as a terminal judgment but as an integral and continuous element throughout the entirety of the educational experience (Zhao & Zhao, 2023).

This broadened perspective necessitates the implementation of assessment practices across all phases of teaching and learning, not merely at its conclusion. With this shift in understanding, the value ascribed to assessment has markedly increased. The quality of assessment conducted throughout the educational process directly influences the quality of learning itself—highlighting its indispensable role in shaping educational outcomes (Chulerk et al., 2025).

In an increasingly globalized world, English proficiency has become a core requirement for educational, professional, and interpersonal success. As a response to this growing demand, the teaching of English as a Foreign Language (EFL) has evolved in both scope and complexity. Central to this evolution is the role of assessment and evaluation, which no longer serve merely as tools for measuring learning outcomes but have become integral to the instructional process itself (Levi & Inbar-Lourie, 2020). In modern pedagogical frameworks, assessment is viewed not only as a summative mechanism used at the end of instruction but also as a formative tool employed continuously to guide learning and inform teaching strategies.

Particularly in EFL contexts, effective assessment practices are essential to ensure that students acquire the four fundamental language skills—listening, speaking, reading, and writing—in an integrated and balanced manner. Imbalances in the development of these skills often result in communicative inadequacies, such

as students' inability to express themselves orally despite strong receptive comprehension. Research shows that bridging the gap between receptive and productive skills requires authentic assessment strategies grounded in real-life language use (Mandasari et al., 2025).

Moreover, internationally recognized frameworks like the Common European Framework of Reference for Languages (CEFR) have provided structured benchmarks for both curriculum design and assessment. These guidelines emphasize the importance of contextually appropriate, level-specific evaluation instruments that measure language proficiency across cognitive, linguistic, and pragmatic dimensions (Kunschak, 2020).

This study critically explores the tools, practices, and underlying principles of assessment and evaluation in EFL classrooms. It aims to bridge theory and practice by examining how language educators apply various assessment strategies to ensure meaningful language acquisition. Particular focus is placed on aligning assessment methods with CEFR standards, integrating skill-based approaches, and distinguishing between summative, formative, and diagnostic purposes in classroom application.

METHOD

This study employed a qualitative content analysis approach based on an extensive review of theoretical literature and practical frameworks surrounding EFL assessment (Patton, 2002). Drawing on official curriculum guidelines, national and international policy documents, and scientific articles from the last decade, the research aimed to synthesize the multidimensional nature of assessment and evaluation practices in English language teaching.

To ensure a comprehensive perspective, data were thematically analyzed through deductive and inductive coding. Deductive codes were drawn from CEFR descriptors and established categories of assessment—formative, summative, diagnostic, and performance-based (Kim, 2021). Inductive codes emerged from close reading of pedagogical research, particularly studies focusing on learner-centered and communicative approaches to evaluation.

The study also examined assessment instruments currently used in EFL classrooms, such as written exams, oral tests, multiple-choice formats, checklists, rubrics, and performance tasks. These tools were evaluated in terms of their alignment with CEFR proficiency levels (A1–C2), suitability for skill integration, and capacity to provide meaningful feedback.

The method was designed to align closely categorized evaluation tools by skill (listening, reading, speaking, writing), followed by specific considerations such as text complexity, item type, objectivity, and student readiness. The data analysis method thus ensured that the results reflect both conceptual rigor and pedagogical relevance.

RESULTS AND DISCUSSION

1. Teaching English as a Foreign Language

The contemporary era is characterized by intensified global interactions, where individuals, driven by various motivations, demonstrate a growing inclination toward learning foreign languages. In alignment with this global phenomenon, the demand for learning English as a foreign language has grown substantially. This rising demand necessitates the development of high-quality educational frameworks for teaching English to non-native speakers, including the implementation of rigorous and effective assessment and evaluation strategies. The foundation of quality education lies in sound decision-making, and such decisions are contingent upon reliable assessment practices (Krolak-Schwerdt et al., 2014).

Language instruction is a comprehensive process aimed at equipping learners with the four essential skills—listening, speaking, reading, and writing—in an integrated and balanced manner. It is not simply an abstract transfer of linguistic knowledge but a holistic endeavor. Prioritizing one skill at the expense of others fragments the learning experience and often results in an incomplete mastery of the language. The frequently expressed sentiment among learners—“I can understand, but I cannot speak”—is a direct reflection of this imbalance and a clear indication of insufficient communicative competence (Coşgun & Hasırcı, 2017).

Consequently, it is crucial in foreign language education to nurture all language skills concurrently and to provide learners with opportunities to apply these skills in authentic, everyday situations. Zulianti et al. (2024) similarly emphasize that for language education to be effective, learners must be able to transform theoretical knowledge into practical usage.

To facilitate this transition from knowledge to practice, language teaching must incorporate materials and activities rooted in real-life contexts. Furthermore, recognizing that language is a living, evolving phenomenon, instruction should present the language in its contemporary and functional forms—forms that address current communicative needs and enable learners to use the language meaningfully in their daily lives. This approach not only reinforces relevance but also counteracts the abstraction that can hinder language acquisition (Nguyen et al., 2023).

As in all fields of education, the successful instruction of English as a foreign language depends on the application of systematic and multidimensional assessment and evaluation procedures (Yulia et al., 2019). Learners are expected to develop proficiency across all four fundamental skills—listening, speaking, reading, and writing. The degree to which learners attain these skills is determined through continuous, formative, and summative assessments.

The instructional process begins with the formulation of clear objectives, followed by the implementation of diverse teaching and learning strategies aimed at achieving these goals. Assessment occurs before instruction (diagnostic), during instruction (formative), and after instruction (summative), providing insights into

learners' preparedness, the effectiveness of instructional delivery, and the overall achievement of learning outcomes (Smolansky et al., 2023).

These goals and outcomes in teaching English as a foreign language are guided by the Common European Framework of Reference for Languages (CEFR), an internationally recognized framework developed by the Council of Europe. The CEFR serves as a comprehensive reference for designing language programs, developing instructional materials, and articulating language proficiency benchmarks across various linguistic contexts (Harsch & Seyferth, 2020).

In the context of English language instruction for non-native speakers, course content and learning objectives are aligned with the CEFR's six proficiency levels—A1, A2, B1, B2, C1, and C2—each defined by specific communicative competencies and descriptors. Since language instruction is not merely about delivering content but also involves interactive feedback and learner-centered assessment, the CEFR also provides guidance on appropriate tools for evaluating linguistic performance in meaningful contexts (Nagai et al., 2020).

Thus, the CEFR plays a pivotal role not only in shaping the curriculum and instructional objectives for teaching English as a foreign language but also in framing assessment strategies that promote comprehensive, functional, and contextually grounded language proficiency.

2. Assessment and Evaluation

Although the terms assessment and evaluation are often used interchangeably, they in fact refer to distinct concepts. The focal point of assessment is observation. Moreover, assessment involves gathering evidence about student learning through systematic observation and translating it into meaningful data (Song & Song, 2023). Similarly, the essence of assessment lies in our ability to observe variations in learners' performance over time (Slamet & Mukminatien, 2024). Both definitions emphasize the centrality of observation. The process of assessment begins with observation, and what compels us to observe is difference. The variation in the degree to which objects possess certain attributes necessitates the act of observation. The process of assessment concludes with the transformation of these observed differences into numerical or symbolic representations. However, assessment alone is insufficient for making judgments. In other words, the act of assessment does not culminate in evaluative judgment. This distinction marks a clear separation between assessment and evaluation.

Evaluation is the process in which the results obtained through assessment are compared against a specific criterion, leading to the formulation of a judgment (Morris et al., 2021). Assessment is thus a necessary precursor to evaluation. Although assessment alone does not suffice for decision-making, it is indispensable within the decision-making process. For example, consider a level placement test in which each item is worth 10 points, and a student correctly answers 5 questions,

earning a score of 50. This numerical result constitutes the assessment process—the student’s knowledge of the subject has been observed and quantified through appropriate instruments. If the passing score for the exam is 70 (the criterion), then the student’s score of 50, when measured against the required 70, results in a judgment—that the student has failed to pass the level. This constitutes the evaluation process.

Assessment and evaluation can be carried out through a variety of methods and instruments. These methods are typically categorized into two broad groups: traditional assessment and evaluation methods (such as written exams, oral exams, multiple-choice questions, true-false questions, matching items, and short-answer questions), and alternative assessment and evaluation methods (such as performance tasks, checklists, and rubrics). Each method has its own benefits and limitations. Therefore, when selecting an assessment method, the primary consideration should be the objective—in other words, the intended learning outcome. Furthermore, it is often advantageous to employ a combination of different methods in order to ensure a comprehensive and purpose-aligned assessment and evaluation process. Therefore, using multiple assessment techniques enables a more accurate representation of learners’ actual language proficiency in EFL context.

3. Assessment Tools Applicable in Teaching English as a Foreign Language

3.1. Written Examinations

In written examinations, learners are expected to reflect on the question, organize their thoughts coherently, and express their answers in written form. The response is composed entirely by the student, without the aid of multiple-choice or predefined answer options (Oktarin et al., 2024). This structure grants examinees a degree of freedom in constructing their responses, thereby encouraging independent expression and linguistic creativity.

While written exams are relatively easy to construct, the process of evaluating them poses considerable challenges, primarily due to the inherently subjective nature of assessment. The difficulty in maintaining objectivity during grading is among the primary limitations of this form of evaluation (Wulyani et al., 2024).

English, as a subject, emphasizes expressive competence. One of the core objectives in teaching English—whether as a first or a foreign language—is to equip learners with the ability to articulate themselves clearly and appropriately in both spoken and written modalities. Hence, written examinations serve not only to measure writing proficiency but also to promote the development of expressive skills in the target language (Cheng & Zhang, 2021).

Importantly, written exams need not be confined solely to writing skills. Integrating productive and receptive skills within the same evaluative framework can foster a more holistic and communicative approach to language learning. For

example, written exams may be used in conjunction with reading comprehension activities. Students can be presented with a reading passage appropriate to their proficiency level—taking into account lexical density, sentence complexity based on the grammar they have learned, and an acceptable number of unfamiliar words—and asked to respond in writing to a series of comprehension questions. In doing so, the assessment targets both reading and writing competencies simultaneously (Namaziandost et al., 2022).

Similarly, listening comprehension—a receptive skill—may also be evaluated through written responses. When employing this approach, the listening text should correspond to the learners' current language level. Ideally, it should utilize grammatical structures previously introduced in instruction, contain only a minimal number of unfamiliar words (typically between three and five), and align with overall learning goals (Hu et al., 2023).

3.2. Oral Examinations

In oral examinations, questions are delivered verbally, and learners are expected to provide spontaneous spoken responses (Davis & Karunathilake, 2005). As in written assessments, the answers in oral exams are not selected from a predefined set but are generated entirely by the examinee. This distinguishes oral assessments from more structured test formats, such as multiple-choice or matching exercises, in which learners merely identify the correct response among given alternatives.

Oral responses, unlike written ones, are expressed through spoken language, requiring real-time linguistic processing and performance. In everyday communication, individuals primarily interact through speaking and listening. Consequently, speaking is a vital skill, both in natural discourse and in second or foreign language acquisition (Istiara et al., 2023).

Speech in real-life contexts often takes the form of either dialogic (interactive) or monologic (independent) communication. Classroom activities designed for teaching English as a foreign language typically reflect these two speech modes, and oral examinations can be structured to evaluate both types of speaking performance (El Shazly, 2021).

The term “examination” has long been associated with anxiety, a reaction that becomes more pronounced in the context of foreign language learning. To mitigate this anxiety, oral assessments should be conducted in environments that approximate authentic communicative settings. The testing atmosphere should be relaxed and conversational, enabling learners to perform as they would in everyday dialogue rather than under the pressure of formal evaluation (Al-Nouh et al., 2015).

Oral examinations are indispensable in assessing both speaking and listening skills, which often operate in tandem during natural language use. However, evaluative subjectivity remains a concern in oral assessments. To enhance reliability

and fairness, it is advisable to complement oral exams with alternative assessment tools—such as standardized rubrics, rating scales, or checklists. These instruments provide structured criteria that support more objective and consistent evaluation, ensuring that the assessment process is as transparent and equitable as possible (Jiang et al., 2021).

3.3. Multiple-Choice Tests

Multiple-choice tests are defined as assessments in which the correct answer is not generated by the respondent, but rather selected from a range of options provided by the test designer(s) (Coşkun et al., 2025). These tests are among the most frequently used instruments in language assessment. However, the exclusive or predominant use of multiple-choice questions in foreign language instruction may lead to certain drawbacks. As learners are not required to produce language themselves, these tests do not engage the productive skills of the language user, thus limiting their ability to demonstrate actual language use.

While it is not inherently problematic to include items that target reading or listening comprehension in a multiple-choice format, the overreliance on such items may stifle creativity and inhibit the development of productive language skills. In this regard, a balanced assessment approach that combines receptive and productive tasks is essential for comprehensive language learning (Cheung et al., 2023).

3.4. Matching Questions

Matching-type questions consist of two parallel lists containing related items that learners are asked to pair appropriately. In the context of teaching English as a foreign language—particularly at the beginner level—matching activities can include tasks such as aligning vocabulary with definitions, words with corresponding images, or antonyms with one another (Juele, 2018). These types of questions can be effectively utilized in both the instruction and assessment of vocabulary, serving as a practical tool for reinforcing word meaning and lexical knowledge.

3.5. True-False Questions

True-false questions are characterized by statements that are either accurate or inaccurate; however, they are not phrased in the form of direct questions, but rather as declarative propositions (D. Brown, 2007). One of the primary limitations of this format is the high probability of guessing correctly—50%—which can negatively impact the reliability of the test. Despite this limitation, including a small number of true-false items in a broader exam structure can introduce variety and reduce monotony, especially when targeting reading or listening comprehension.

These items are well-suited to assessing factual recall and basic understanding of read or heard content. However, they are generally unsuitable for measuring higher-order cognitive skills, such as analysis, synthesis, or evaluation, which require more complex responses and deeper engagement with the material.

3.6. Short-Answer Questions

Similar to essay-type questions, short-answer questions require written responses. However, unlike longer written responses that may span multiple sentences or paragraphs, answers to short-answer items are brief—often limited to a single word, phrase, sentence, or numerical value (Brenner et al., 2024). These items can take the form of direct questions or incomplete statements, requiring the respondent to supply the missing element.

Short-answer questions are commonly employed in teaching English as a foreign language, particularly within reading comprehension activities. For example, learners might be presented with a text and subsequently asked questions that require brief written answers. These questions are also suitable for tasks involving everyday expressions or introductory dialogues, especially at elementary levels. In such cases, students may be asked to complete missing parts of dialogues, thereby reinforcing familiar patterns and structures (Nguyentan et al., 2022).

While short-answer items are frequently used in the assessment of reading comprehension, employing them to evaluate writing proficiency may not yield accurate results. Since writing is an inherently creative and extended process, tasks that demand only a few words or a single sentence fall short of adequately capturing a learner's writing ability.

3.7. Performance-Based Tasks

Performance assessment aims to evaluate a learner's capacity to *do* or perform a task. It emphasizes the application of knowledge in practice, often requiring learners to engage in hands-on activities either individually or collaboratively (Trautmann et al., 2024).

In the context of teaching English as a foreign language, performance tasks are especially effective in the area of prepared speech. Learners can be assigned specific tasks ahead of time, allowing them to plan and rehearse before performing. For example, students might be grouped in pairs and assigned everyday scenarios to act out after a period of preparation. One student might take on the role of a vendor and the other as a customer at a marketplace; alternatively, a pair may role-play as a waiter and a diner. Enhancing such tasks with preliminary activities—like preparing a sample menu—can render the experience more authentic and engaging.

By integrating performance-based activities, educators can foster a more dynamic, communicative learning environment that supports both linguistic competence and real-world language use.

3.8. Checklist

The checklist serves as a supportive tool designed to document students' performance . It is used to observe learners as they engage in a behavior or produce a product, enabling the evaluator to determine whether the expected behaviors have been demonstrated (Walvoord et al., 2023).

In the context of teaching English as a foreign language, specific behaviors are anticipated across all proficiency levels (A1, A2, B1, B2, C1, C2) for each of the four core skills: listening, reading, speaking, and writing. These expected behaviors can be listed and used as observational criteria to determine whether students exhibit the required competencies. Consequently, the evaluator can identify which targeted behaviors have been successfully acquired and which have not, even though they were intended learning outcomes. This facilitates the detection and remediation of learning gaps (Uchida & Negishi, 2025).

Evaluating language performance without the aid of any measurement tool risks overlooking numerous critical elements during the assessment process. By contrast, employing a checklist that clearly outlines the desired behaviors offers significant support in determining whether those behaviors have been realized (Hastomo et al., 2024).

3.9. Analytic Rubric

An analytic rubric is a scoring guide that includes well-defined criteria for describing performance. It is a tool developed to evaluate the components of a specific task (Rock, 2022). Unlike simple checklists that merely record whether a behavior has occurred, an analytic rubric provides graded levels of performance for each criterion.

Students can use analytic rubrics as a roadmap before performing a task or producing a product. In this way, the rubric not only supports assessment but also serves an instructional function. In the teaching of English as a foreign language, analytic rubrics are particularly beneficial for assessing productive skills such as speaking and writing .(Imbler et al., 2023)

Exposure to the rubric's categories prior to the task can provide learners with a clearer understanding of the expectations, increasing the likelihood of focusing on key aspects of performance. For the evaluator, an analytic rubric eliminates ambiguity by specifying the level of quality associated with different score points.

4. Assessment and Evaluation as a Foreign Language

The primary aim of teaching English as a foreign language is to enable learners to use the language proficiently and in a balanced manner across the four foundational skills: listening, reading, speaking, and writing. For each proficiency level, there are specific learning outcomes expected of students. Assessment and evaluation practices are essential tools in facilitating learners' progress toward

these outcomes and in determining their achievement levels (Slamet & Mukminatien, 2024).

Assessment activities conducted before the instructional process serve to place learners in appropriate language classes. During instruction, formative assessment helps identify and address learning deficiencies. Summative assessment at the end of the process is crucial for evaluating learning outcomes. Assessment tasks aimed at evaluating learners' language skills are structured according to the four core competencies (Brown, 2004).

4.1. Assessment and Evaluation of Listening Skills

Listening, which plays a critical role in daily communication and holds a prominent place in foreign language instruction. In the teaching of English as a foreign language, students are expected to comprehend listening texts appropriate to their proficiency level. The nature of expected comprehension varies by level. For instance, at the basic level, learners are expected to understand the main ideas of a listening passage, whereas at advanced levels, they are expected to infer implicit meanings (Şendağ et al., 2018).

Both the complexity of the text and the expected level of understanding increase with proficiency. Listening passages for lower levels should be shorter and simpler, increasing in length and complexity as learners advance. It is important to note that the type of listening text used in assessment significantly affects comprehension. Understanding a narrative text requires a different level of cognitive effort and attentiveness compared to processing an expository text. Since expository texts demand more intense mental engagement, it is acceptable to select shorter texts (fewer words) than the prescribed range when using such texts for listening assessment (Nation, 2006).

Another critical feature of listening texts is the accuracy of their audio delivery. Proper stress, intonation, and pronunciation are essential because language learners tend to model what they hear. Additionally, the number of unfamiliar words should be limited to avoid distracting learners and disrupting comprehension (Partiwi, 2022).

Listening skills can be evaluated using both traditional and alternative assessment tools. self-assessment, peer assessment, observation forms, and checklists can be used to assess observable and physical aspects of listening skills. Moreover, comprehension can be measured through true/false questions, short-answer items, and matching exercises (Brown, 2004).

4.2. Measuring and Assessing Reading Skills

The concept of reading has been defined as “a complex cognitive process involving decoding symbols to derive meaning” and as “an active process that integrates word recognition, comprehension, fluency, and motivation” (Anggeraini

et al., 2020). Similar to listening, reading is a multifaceted cognitive skill requiring various mental activities.

In teaching English as a foreign language, it is essential that texts used to measure reading skills are appropriate in terms of content and length relative to learners' proficiency levels. For example, texts for beginner readers should be "short, contextually simple, and limited in new vocabulary to avoid cognitive overload." Assessing an A1 learner with a long, complex text would be counterproductive. Moreover, texts should align with the grammatical structures and vocabulary covered at the learner's level, with adaptations as necessary (Perfetti et al., 2008). Similar to listening materials, reading texts should limit unfamiliar words to support comprehension and motivation.

Various question types such as multiple-choice, true/false, short-answer, and matching exercises are effective for assessing reading comprehension (Oakhill et al., 2014). At advanced levels, reading may be integrated with productive tasks such as oral discussions or written summaries based on the reading material.

4.3. Measuring and Assessing Speaking Skills

Speaking is a primary mode of human communication, essential in daily interaction and foreign language learning. Since language primarily functions as a communication tool, developing functional speaking skills is critical for learners at all proficiency levels (Kusuma, 2022).

Speaking skills assessment can be categorized into interactive conversation and individual self-expression. Creating assessment scenarios that reflect both categories provides a balanced evaluation. Topics should be functional and relevant to everyday life to promote authentic communication. Reducing anxiety during speaking assessments is important because speaking is spontaneous and learners often feel pressure (Ma et al., 2021).

Performance-based assessments, such as checklists and rating scales, aligned with proficiency objectives, are valuable tools. Employing multiple raters can enhance reliability. Peer assessment is beneficial if criteria are clear and feedback is constructive to prevent discouragement (Brown, 2004).

4.4. Assessing Writing Skills

Writing, unlike speaking, allows for planning, revising, and refining output, making it a reflective process (Hyland, 2003). It is often the last skill to develop due to less frequent daily use and limited practice at basic proficiency levels. However, balanced language learning requires adequate emphasis on writing.

Linking writing tasks with receptive skills such as reading and listening can facilitate writing development. For instance, learners may be asked to complete unfinished texts or write responses to listening prompts, encouraging integrated skills development (Yu et al., 2022).

Writing assessment can involve peer and self-assessment, rating scales, and checklists tailored to proficiency and writing goal. Because written work is tangible, it enables detailed feedback and multiple drafts, which are crucial for writing improvement (Acar, 2023). Revisions following feedback help learners internalize writing conventions and improve performance.

CONCLUSION

In recent years, there has been a growing demand for learning English as a foreign language. This increasing demand has brought the significance of planning, instructional processes, and assessment and evaluation in English language teaching to the forefront. Achieving the goals of language instruction requires careful planning, sound decision-making, and proper structuring of the teaching and learning process. Assessment and evaluation constitute a prerequisite for making informed decisions.

Assessment practices are conducted to identify students and place them in appropriate classes prior to the instructional process, to monitor progress and address learning gaps during the process, and to determine learners' achievement levels upon its completion. Continuous assessment and evaluation during English as a foreign language instruction facilitate the attainment of instructional objectives and provide empirical evidence of student success.

The overarching aim in teaching English as a foreign language is to develop the four fundamental language skills in a balanced manner. These skills are grouped into receptive (comprehension) and productive (expression) categories. Listening and reading constitute comprehension skills, whereas speaking and writing form expression skills. It is beneficial to teach these skills in an integrated manner whenever possible, as multiple language skills are often employed simultaneously in daily communication. For example, during a conversation, speakers alternate between speaking and listening. Similarly, while attending a lecture or conference, one listens to the speaker and simultaneously takes notes, thereby engaging writing skills. Therefore, teaching and assessing language skills in an interconnected manner is advantageous. Instead of evaluating listening, reading, speaking, and writing skills independently, designing assessment tasks that integrate at least two skills can yield more meaningful results.

Assessment instruments used in English language teaching must be carefully selected to align with the proficiency levels targeted by the learning outcomes. Employing multiple appropriate assessment tools for each skill can enhance the variety and robustness of the evaluation process. For listening and reading comprehension, short-answer questions, true/false items, multiple-choice questions, and open-ended questions requiring extended responses can be utilized. Because speaking is an instantaneous performance skill, assessment of speaking behaviors can be facilitated by the use of rating scales, which clearly define the expected behaviors and ease the evaluator's task. Additionally, checklists, peer

assessments, oral examinations, and performance tasks may be employed as complementary evaluation tools.

Writing skill assessment benefits from the tangible nature of the produced texts, which can be reviewed repeatedly. This allows iterative evaluation and rewriting, supporting progressive skill development. Encouraging learners to self-assess their written work fosters metacognitive awareness and autonomy. Peer assessment can also be incorporated by enabling students to evaluate each other's texts. Using checklists aligned with targeted learning outcomes helps identify which objectives have been met and which require further attention, allowing for targeted remediation.

REFERENCES

- Acar, A. S. (2023). Genre pedagogy: A writing pedagogy to help L2 writing instructors enact their classroom writing assessment literacy and feedback literacy. *Assessing Writing*, 56, 100717. <https://doi.org/10.1016/j.asw.2023.100717>
- Al-Nouh, N. A., Abdul-Kareem, M. M., & Taqi, H. A. (2015). EFL College Students' Perceptions of the Difficulties in Oral Presentation as a Form of Assessment. *International Journal of Higher Education*, 4(1). <https://doi.org/10.5430/ijhe.v4n1p136>
- Anggeraini, Y., Nurhasanah, & Madenta, T. (2020). EFL Learners' Reading Habit and Their Reading Comprehension Enhancement through Partner Reading. *Getsempena English Education Journal*, 7(2), 294–306. <https://doi.org/10.46244/GEEJ.V7I2.985>
- Brenner, J. M., Fulton, T. B., Kruidering, M., Bird, J. B., Willey, J., Qua, K., & Olvet, D. M. (2024). What have we learned about constructed response short-answer questions from students and faculty? A multi-institutional study. *Medical Teacher*, 46(3), 349–358. <https://doi.org/10.1080/0142159X.2023.2249209>
- Brown, D. (2007). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Pearson Education.
- Brown, D. H. (2004). *Language Assessment - Principles and Classroom Practice*. San Francisco State University. https://www.academia.edu/26575645/H_Douglas_Brown_Language_Assessment_Principles_and_Classroom_Practice
- Cheng, X., & Zhang, L. J. (2021). Teacher written feedback on English as a foreign language learners' writing: Examining native and nonnative English-speaking teachers' practices in feedback provision. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.629921>
- Cheung, B. H. H., Lau, G. K. K., Wong, G. T. C., Lee, E. Y. P., Kulkarni, D., Seow, C. S., Wong, R., & Co, M. T.-H. (2023). ChatGPT versus human in generating medical graduate exam multiple choice questions—A multinational prospective study

- (Hong Kong S.A.R., Singapore, Ireland, and the United Kingdom). *PLOS ONE*, 18(8), e0290691. <https://doi.org/10.1371/journal.pone.0290691>
- Chulerk, P., Duangchinda Sathianpan, S., & Khlaisan, J. (2025). Mobile Assisted Language Learning as a Formative Assessment Method to Enhance the English Competence of EFL Learners to the CEFR B1 Level. *REFlections*, 32(1), 462–486. <https://doi.org/10.61508/refl.v32i1.280328>
- Coşgun, G., & Hasırcı, B. (2017). The impact of English medium instruction (EMI) on students' language abilities. *International Journal of Curriculum and Instruction*. <https://eresearch.ozyegin.edu.tr/handle/10679/5727>
- Coşkun, Ö., Kiyak, Y. S., & Budakoğlu, I. İ. (2025). ChatGPT to generate clinical vignettes for teaching and multiple-choice questions for assessment: A randomized controlled experiment. *Medical Teacher*, 47(2), 268–274. <https://doi.org/10.1080/0142159X.2024.2327477>
- Davis, M. H., & Karunathilake, I. (2005). The place of the oral examination in today's assessment systems. *Medical Teacher*, 27(4), 294–297. <https://doi.org/10.1080/01421590500126437>
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(3), e12667. <https://doi.org/10.1111/EXSY.12667>
- Harsch, C., & Seyferth, S. (2020). Marrying achievement with proficiency – Developing and validating a local CEFR-based writing checklist. *Assessing Writing*, 43, 100433. <https://doi.org/10.1016/j.asw.2019.100433>
- Hastomo, T., Mandasari, B., & Widiati, U. (2024). Scrutinizing Indonesian pre-service teachers' technological knowledge in utilizing AI-powered tools. *Journal of Education and Learning (EduLearn)*, 18(4), 1572–1581. <https://doi.org/10.11591/edulearn.v18i4.21644>
- Hu, P., Gong, Y., Lu, Y., & Ding, A. W. (2023). Speaking vs. listening? Balance conversation attributes of voice assistants for better voice marketing. *International Journal of Research in Marketing*, 40(1), 109–127. <https://doi.org/10.1016/j.ijresmar.2022.04.006>
- Hyland, K. (2003). *Second language writing*. Cambridge University Press.
- Imbler, A. C., Clark, S. K., Young, T. A., & Feinauer, E. (2023). Teaching second-grade students to write science expository text: Does a holistic or analytic rubric provide more meaningful results? *Assessing Writing*, 55, 100676. <https://doi.org/10.1016/j.asw.2022.100676>
- Istiara, F., Hastomo, T., & Indriyanta, W. A. (2023). A study of students' engagement and students' speaking skill: A correlational research. *TEKNOSASTIK*, 21(1), 1–7. <https://doi.org/10.33365/TS.V21I1.2198>
- Jiang, Jong, M. S., Lau, W. W., Chai, C. S., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a

- flipped classroom. *Australasian Journal of Educational Technology*, 37(2), 110–131.
- Juele, L. (2018). Authentic assessments: A critical thinking and engagement tool for online courses. In T. Bastiaens, M. Van Braak, & L. Brown (Eds.), *Proceedings of EdMedia: World Conference on Educational Media and Technology* (pp. 1967–1981). Association for the Advancement of Computing in Education (AACE). <https://www.learntechlib.org/primary/p/184435/>
- Kim, S. (2021). Generalizability of CEFR Criterial Grammatical Features in a Korean EFL Corpus across A1, A2, B1, and B2 Levels. *Language Assessment Quarterly*, 18(3), 273–295. <https://doi.org/10.1080/15434303.2020.1855647>
- Krolak-Schwerdt, S., Glock, S., & Böhmer, M. (2014). *Teacher's professional development assessment, training, and Learning*. Springer Science & Business Media.
- Kunschak, C. (2020). CEFR, CLIL, LOA, and TBLT – Synergising Goals, Methods and Assessment to Optimise Active Student Learning. In *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms* (pp. 85–108). Springer International Publishing. https://doi.org/10.1007/978-3-030-54128-6_4
- Kusuma, I. P. I. (2022). EFL preservice teachers' technology integration in managing and teaching speaking skills during emergency remote teaching. *Profile: Issues in Teachers' Professional Development*, 149–165. <https://doi.org/https://doi.org/10.15446/profile.v24n2.97497>
- 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>
- Ma, M., Wang, C., & Teng, M. F. (2021). Using learning-oriented online assessment to foster students' feedback literacy in L2 writing during COVID-19 pandemic: A case of misalignment between micro- and macro- contexts. *The Asia-Pacific Education Researcher*, 30(6), 597–609. <https://doi.org/10.1007/s40299-021-00600-x>
- Mandasari, B., Basthomi, Y., Hastomo, T., Afrianto, Hamzah, I., & Aminatun, D. (2025). The Snapshots of Indonesian Pre-Service English Teachers' Perspectives on Integrating Technology-Based Tools to Rural Schools. *Voices of English Language Education Society*, 9(1), 42–57. <https://doi.org/10.29408/veles.v9i1.27965>
- Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9(3). <https://doi.org/10.1002/rev3.3292>
- Nagai, N., Birch, G. C., Bower, J. V., & Schmidt, M. G. (2020). *Integrating Learning, Teaching, and Assessment* (pp. 197–240). https://doi.org/10.1007/978-981-15-5894-8_5

- Namaziandost, E., Razmi, M. H., Ahmad Tilwani, S., & Pourhosein Gilakjani, A. (2022). The impact of authentic materials on reading comprehension, motivation, and anxiety among Iranian male EFL learners. *Reading & Writing Quarterly*, 38(1), 1–18. <https://doi.org/10.1080/10573569.2021.1892001>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? . *Canadian Modern Language Review*, 63(1), 59–82.
- Nguyen, N. T., Baker, A., Eady, M., & Wright, J. (2023). Vietnamese EFL Novice Teachers' Pedagogical Decisions within a Mandated Communicative Language Curriculum. *The Journal of AsiaTEFL*, 20(2), 338–356. <https://doi.org/10.18823/asiatefl.2023.20.2.7.338>
- Nguyentan, D.-C., Gruenberg, K., & Shin, J. (2022). Should multiple-choice questions get the SAQ? Development of a short-answer question writing rubric. *Currents in Pharmacy Teaching and Learning*, 14(5), 591–596. <https://doi.org/10.1016/j.cptl.2022.04.004>
- Oakhill, J., Cain, K., & Elbro, C. (2014). Understanding and teaching reading comprehension: A handbook. *Understanding and Teaching Reading Comprehension: A Handbook*, 1–129. <https://doi.org/10.4324/9781315756042/UNDERSTANDING-TEACHING-READING-COMPREHENSION-JANE-OAKHILL-KATE-CAIN-CARSTEN-ELBRO>
- O'Dwyer, F., & de Boer, M. (2015). Approaches to assessment in CLIL classrooms: Two case studies. *Language Learning in Higher Education*, 5(2), 397–421. <https://doi.org/10.1515/cercles-2015-0019>
- Oktarin, I. B., Saputri, M. E. E., Magdalena, B., Hastomo, T., & Maximilian, A. (2024). Leveraging ChatGPT to enhance students' writing skills, engagement, and feedback literacy. *Edelweiss Applied Science and Technology*, 8(4), 2306–2319. <https://doi.org/10.55214/25768484.v8i4.1600>
- Partiwi, S. (2022). The Use of Podcasts for Listening Comprehension. *Journal of English Language and Education*, 7(2), 77–83. <https://doi.org/10.31004/jele.v7i2.274>
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Sage Publications.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2008). The Acquisition of Reading Comprehension Skill. *The Science of Reading: A Handbook*, 227–247. <https://doi.org/10.1002/9780470757642.CH13>
- Rock, K. (2022). Constructing a data-based analytic rubric for an academic blog post. *Assessing Writing*, 51, 100602. <https://doi.org/10.1016/j.asw.2021.100602>
- Şendağ, S., Gedik, N., & Toker, S. (2018). Impact of repetitive listening, listening-aid and podcast length on EFL podcast listening. *Computers & Education*, 125, 273–283. <https://doi.org/10.1016/J.COMPEDU.2018.06.019>
- Slamet, J., & Mukminatien, N. (2024). Developing an Online Formative Assessment Instrument for Listening Skill through LMS. *LEARN Journal*, 17(1), 188–211.

- Smolansky, A., Cram, A., Radulescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education. *Proceedings of the Tenth ACM Conference on Learning @ Scale*, 378–382. <https://doi.org/10.1145/3573051.3596191>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Trautmann, S. T., Vollmann, M., & Becker, C. (2024). Performance prediction and performance-based task allocation. *Journal of Economic Behavior & Organization*, 220, 354–368. <https://doi.org/10.1016/j.jebo.2024.02.027>
- Uchida, S., & Negishi, M. (2025). Assigning CEFR-J levels to English learners' writing: An approach using lexical metrics and generative AI. *Research Methods in Applied Linguistics*, 4(2), 100199. <https://doi.org/10.1016/j.rmal.2025.100199>
- Walvoord, B. E., Stevens, D. D., & Levi, A. J. (2023). *Introduction to Rubrics*. Routledge. <https://doi.org/10.4324/9781003445432>
- Wulyani, A. N., Widiati, U., Muniroh, S., Rachmadhany, C. D., Nurlaila, N., Hanifiyah, L., & Sharif, T. I. S. T. (2024). Patterns of utilizing AI-assisted tools among EFL students: Need surveys for assessment model development. *LLT Journal: A Journal on Language and Language Teaching*, 27(1), 157–173. <https://doi.org/10.24071/llt.v27i1.7966>
- Yu, S., Di Zhang, E., & Liu, C. (2022). Assessing L2 student writing feedback literacy: A scale development and validation study. *Assessing Writing*, 53(7), 100643. <https://doi.org/10.1016/j.asw.2022.100643>
- Yüce, E., & Mirici, İ. H. (2023). Self-assessment in EFL classes of secondary education in Türkiye: The Common European Framework of Reference for Languages (CEFR)-based implementations. *Pegem Journal of Education and Instruction*, 13(1), 349–359. <https://doi.org/10.47750/pegegog.13.01.38>
- Yulia, A., Husin, N. A., Anuar, F. I., & Alam, S. (2019). Channeling assessments in English language learning via interactive online platforms. *Studies in English Language and Education*, 6(2), 228–238. <https://doi.org/10.24815/SIELE.V6I2.14103>
- Zhao, H., & Zhao, B. (2023). Co-constructing the assessment criteria for EFL writing by instructors and students: A participative approach to constructively aligning the CEFR, curricula, teaching and learning. *Language Teaching Research*, 27(3), 765–793. <https://doi.org/10.1177/1362168820948458>
- Zulianti, H., Hastuti, H., Nurchurifiani, E., Hastomo, T., Maximilian, A., & Ajeng, G. D. (2024). Enhancing Novice EFL Teachers' Competency in AI-Powered Tools Through a TPACK-Based Professional Development Program. *World Journal of English Language*, 15(3), 117. <https://doi.org/10.5430/wjel.v15n3p117>

