

Chapter 15

Designing Authentic English Language Assessment: Principles, Practices, and Innovations for Meaningful Language Learning

Siti Mariam

Universitas Islam Negeri Walisongo Semarang

sitimariam@walisongo.ac.id

A. THE CONTEMPORARY CHALLENGE

It is becoming more and more common for English language assessments to go beyond measuring discrete linguistic knowledge to examine students' ability to use the language in ways that are appropriate, strategic, and meaningful in situations that are similar to real-world communication. In order to meet this demand, authentic assessment creates assignments that call on students to use their language skills, communication techniques, higher-order thinking, and contextual awareness in order to complete meaningful tasks. According to recent research, authenticity is more than just employing "real-life" materials; it also refers to how well assessment tasks, situations, cognitive processes, and expected performances match language use outside of the classroom (Hasrol, S. B., et al, 2022; Mariam, S., et al, 2025b).

However, a comprehensive analysis of 96 studies revealed that authenticity in second-language assessment is still conceptually and

methodologically underdeveloped, especially when it comes to defining distinct connections between assessment tasks and target-language-use domains. With the growing emphasis on communicative competence, learner agency, multimodality, and digital literacy in language education, there is an urgent need to rethink authentic English assessment. In order to provide more comprehensive proof of learners' functional language skills, authentic assessment can include activities like presentations, group problem-solving, interviews, portfolios, simulations, projects, reflective writing, and real-world conversation. According to recent research, authenticity is further conceptualized as a continuum that connects classroom instruction with contextually placed and spontaneous L2 engagement, including mixed and asynchronous settings (Kofinas, A. K., et al, 2025; Quinlan, K. M., et al, 2024).

At the same time, traditional beliefs regarding assessment validity, authorship, fairness, and academic integrity are being called into question by the quick development of generative artificial intelligence (GenAI). When AI is integrated into language assessment ecosystems, it is crucial to re-evaluate reliability, validity, fairness, practicality, and sustainability, according to recent reviews (Zhang, T., et al, 2025). In order to connect assessment activities with meaningful language use while maintaining reliable evidence of individual learning, genuine English language assessment design necessitates a validity-driven, learner-centred, context-sensitive, and ethically informed approach. This viewpoint offers the basis for rethinking authentic assessment as an essential part of modern English language instruction rather than just an alternate testing method.

B. THE PRACTICAL SOLUTIONS

In order to create real English language assessments, it is necessary to move away from evaluating students' knowledge of the language and toward assessing their ability to use it in contextualized, meaningful, and intentional contexts. However, authenticity should not be limited to putting students in "real-life" scenarios. According to (Fawns, T., et al, 2025; Hasrol, S. B., et al, 2022) research makes a distinction between situational authenticity, which is the degree to

which assessment tasks and target-language-use contexts are similar, and interactional authenticity, which is the degree to which learners use cognitive and communicative processes that are similar to those used outside of assessment settings. Thus, comprehensive alignment between learning aims, target-language activities, assessment tasks, performance standards, learner agency, and feedback is necessary for practical implementation. The following ideas offer a framework for creating English language assessments that are more inclusive, valid, and instructional.

Begin with Real-World Language-Use Needs

Determining where, why, with whom, and how students are expected to utilize English is the first workable option. Before creating tasks, assessment designers should perform a target-language-use study. Participating in academic conversations, presenting research, writing professional emails, explaining information to non-specialists, working on projects, conducting interviews, or interacting via digital platforms are some examples of relevant circumstances for university EFL learners.

This method stops teachers from merely designating traditional assessments as "authentic." Numerous studies on authenticity, according to (Ajjawi, R., et al, 2024; Chuang, P.-L., & Yan, 2025) were unable to establish sufficiently explicit linkages between assessment tasks and target-language-use domains. As a result, they can develop an authenticity mapping matrix with four components: (1) real-world communication scenario; (2) language function; (3) evaluation task; and (4) observable performance. For instance, the evaluation might ask students to answer a client question, take part in a simulated meeting, and write a follow-up email if they require English for professional communication.

Replace Isolated Test Items with Performance-Based Tasks

Learners should have the chance to demonstrate integrated language performance during authentic assessment. Teachers can use presentations, simulations, portfolios, projects, interviews, debates,

problem-solving exercises, podcasts, role plays, and group projects in addition to multiple-choice vocabulary, grammar, and reading problems.

These activities allow for the simultaneous evaluation of several aspects of communicative skill. Pronunciation, vocabulary, grammatical control, fluency, organization, audience awareness, and strategic communication can all be demonstrated by a student creating a five-minute podcast, for example. Collaboration and critical thinking can also be evaluated through a problem-based group exercise (Hasbi, 2026). In addition to highlighting the importance of teacher training and institutional support, a recent systematic evaluation of authentic assessment in higher education found significant correlations between authentic assessment and 21st-century abilities like critical thinking, problem-solving, and teamwork. Performance tasks should therefore be chosen because they produce reliable proof of the desired learning outcomes rather than because they seem novel.

Integrate Authenticity with Constructive Alignment

Authentic assessment should be systematically aligned with learning outcomes, classroom activities, and assessment criteria. A practical design sequence is:

Learning outcome → real-world language practice → authentic task → evidence → criteria → feedback.

For instance, students may be required to "communicate recommendations clearly and persuasively in English" as part of a learning objective. Students may be required to submit a suggestion for enhancing a university service to a fictitious institutional committee as part of the relevant assessment. Clarity of argument, utilization of supporting evidence, communicative efficacy, language control, audience awareness, and responsiveness to inquiries are a few examples of assessment criteria. This congruence is crucial since validity cannot be ensured by authenticity alone (Aryadoust, 2023; Mariam, S., et al, 2025a) warn against trusting that an apparently realistic assignment will inevitably yield correct interpretations of learner capacity and highlights the ongoing significance of validity

issues in language assessment. Therefore, authenticity should not be used as a stand-alone quality mark but rather as a part of a larger validity argument.

Develop Transparent Analytic Rubrics

Authentic performance requires assessment criteria that make expectations visible to learners. Analytic rubrics are particularly useful because they separate different dimensions of performance and provide more informative feedback than a single holistic score.

For English language assessment, a rubric may include:

Dimension	Indicators
Communicative effectiveness	Message is clear and appropriate for the audience
Language use	Vocabulary and grammar support meaning
Fluency	Communication proceeds with appropriate flow
Organization	Ideas are logically structured
Interaction	Learner responds appropriately to others
Critical thinking	Claims are supported by relevant reasoning/evidence
Contextual appropriateness	Language choices suit the situation

Rubrics should avoid overemphasizing grammatical accuracy when the purpose is communicative performance. They should instead distinguish between errors that genuinely interfere with meaning and minor deviations that do not. This is particularly relevant in multilingual and Global Englishes contexts, where successful communication cannot always be equated with conformity to a single native-speaker norm.

Incorporate Learner Choice and Agency

When students have some choice over what they generate, how they demonstrate learning, or which topics they cover, authentic assessment takes on greater significance. As long as each choice produces equivalent proof of the targeted learning goals, learners might choose to produce a podcast, video presentation, written proposal, infographic with an oral explanation, or digital portfolio. According to recent studies on educational authenticity, student choice can also play a significant role in authentic assessment, indicating that meaningful engagement is not only produced by real-world linkages. As a result, rather than offering unrestricted choice, educators can offer constrained flexibility. For instance, students can select the subject and media while adhering to standard standards for audience awareness, organization, language use, and communicative effectiveness.

Use Portfolios for Longitudinal Evidence

There is little evidence of language progress from a single performance. A better option is provided by portfolio assessment, which tracks students' development throughout a variety of assignments. First drafts, updated writing, taped presentations, reflective remarks, peer and instructor criticism, and finished items can all be included in a portfolio. Because they enable students to show growth over time and link assessment and reflection, digital portfolios are especially helpful. Teachers can inquire, "What have you learned, what has improved, and what evidence demonstrates that improvement?" instead of, "What score did you obtain?" As a result, evaluation becomes a learning process rather than a final measurement task (). Because writing, speaking, project-based learning, academic English, and professional communication require iterative progress rather than one-time performance, portfolio evaluation is particularly suitable for these disciplines.

Integrate Formative Assessment and Feedback

Authentic assessment should not be limited to final grading. It should incorporate drafting, feedback, revision, reflection, and resubmission. A practical cycle is:

*Task briefing → initial performance → peer/teacher
feedback → revision → final performance → reflection.*

Students creating an academic presentation, for instance, could submit an outline, obtain input from peers, record an initial presentation, get criticism from teachers, make revisions, and then submit the final version along with a brief remark. Because students utilize assessment data to enhance subsequent performance, this method raises the assessment's learning value. It also mirrors communication in the real world, where projects, presentations, and professional writing are often edited before being published or delivered.

Embed Multimodal and Digital Assessment

Multimodal settings, such as video conferencing, social media, podcasts, digital presentations, collaborative documents, and online discussion forums, are increasingly used in modern English communication. Therefore, authentic assessment ought to go beyond conventional written exams. Digital storytelling, video essays, podcasts, online presentations, collaborative blogs, e-portfolios, and simulated professional interactions are all ways that educators might evaluate their students. Crucially, technology should support the assessment framework rather than create new challenges. The desired English-learning objectives should be the main criteria used to evaluate students, not their proficiency with complex software. Additionally, technology can increase the possibilities for real audience engagement. For instance, students can make an educational podcast for a specific audience, perform a simulated online interview, or produce an English-language informational video for potential students.

Design Inclusive and Accessible Assessment

Accessibility, equity, and authenticity should all be united. If a task adds needless obstacles unrelated to the construct being evaluated, it may be very realistic but unjust. Thus, accessibility, language proficiency, disability, technology access, and various forms of communication should all be taken into account by educators. It is possible to use a variety of representation and expression methods without lowering academic standards. For instance, while upholding comparable performance standards, students may exhibit spoken communication through live or recorded presentations as needed. Relevant barriers can be further reduced with the use of accessible instructions, flexible submission formats, subtitles, transcripts, screen-reader-compatible materials, and sufficient preparation time. This idea is in line with current criticisms of authentic evaluation, which stress that authenticity shouldn't always be viewed as intrinsically advantageous or inclusive (Fawns, T., et al, 2025; Mariam, S., et al, 2025) advocate for a more critical view of authenticity and warn against presuming that authentic assessment would automatically address difficult educational issues like inclusiveness or future career preparedness.

Establish Robust Validity and Reliability Procedures

Teachers require systems to guaranty reliable assessment decisions because authentic activities are frequently complex and open-ended. Clear scoring criteria, rater training, moderation, exemplars, double marking for specific performances, and a methodical examination of assessment outcomes are examples of practical solutions. As benchmark examples, teachers can gather samples of high-, medium-, and low-level performances. Moderation meetings can assist in establishing uniform interpretations of rubric descriptors when numerous assessors are participating. However, authenticity should not be sacrificed in the name of dependability. The goal is to ensure that many assessors draw justifiable conclusions regarding comparable performances rather than to artificially standardize complex communication.

Reconsider Assessment in the Age of Generative AI

Authentic evaluation is necessary for generative AI to become more evidence-rich and process-oriented. When students can get significant help from AI systems, traditional take-home essays could not offer enough proof of individual learning. Teachers might rethink evaluation around procedures that are more challenging to fully outsource rather than just barring AI. Oral defences, staged assignments, annotated drafts, reflective commentary, in-class components, individualized tasks, source evaluation, group problem-solving, and brief interviews regarding the student's work are all helpful tactics. Additionally, students may be required to record their use of AI, critically assess language produced by AI, and explain which recommendations they approved or disapproved of.

Validity, reliability, fairness, practicality, and sustainability are key issues in GenAI-supported language evaluation, according to recent systematic research. Similar to this, Zhang, T., et al, (2025) stress that AI can both challenge preconceived notions about evaluation and learner performance and generate opportunities. AI-aware authenticity, which assesses genuine human decision-making, communication, thought, and critical involvement rather than just producing polished writing, should be the aim (Fajaria et al., 2026).

Implement Authentic Assessment Gradually

Lastly, schools should refrain from instantly substituting authentic assessment for all traditional exams. A significant amount of teaching time, technology, professional development, and moderation may be needed for authentic tasks. The 2024 systematic evaluation of authentic assessment identifies resource limitations, training needs, and resistance as major implementation obstacles. Starting with one authentic assignment per course, assessing its efficacy, gathering teacher and student input, updating the rubric, and progressively broadening the approach are all practical implementation strategies. Teachers can receive assistance from professional learning communities in co-creating assignments and overseeing student achievement. Additionally, evaluation policies that acknowledge

digital evidence, authentic performance, accessibility, and ethical AI use should be developed by institutions.

C. THE FUTURE DIRECTION

Instead of isolated classroom assessments, the future of authentic English language assessment should shift toward dynamic, multimodal, learner-centred, and AI-aware assessment ecosystems. First, by investigating both situational authenticity the correspondence between assessment and target-language-use contexts and interactional authenticity the cognitive and communicative processes involved in task completion. Future research should strengthen conceptual and empirical models of authenticity (Ajjawi, R., et al, 2024; Hasrol, S. B., et al, 2022). Second, learner agency and meaningful choice should be incorporated more and more into authentic assessment. According to recent study, students' views of authenticity are impacted by possibilities for choice and meaningful engagement with assessment activities in addition to real-world applicability. Therefore, while adhering to standard validity standards, future assessment designs should let students choose subjects, audiences, genres, or communication styles. Third, authenticity will need to be fundamentally re-evaluated in light of generative AI. Because AI can produce persuasive answers to a variety of challenging problems, authentic evaluation by itself cannot ensure academic integrity (Kofinas, A. K., et al, 2025).

As a result, rather than focusing mostly on final written results, future assessments should highlight learning processes, oral engagement, reflection, performance, evidence of development, and transparent AI use. In the end, authentic, inclusive, technologically responsive, and validity-driven assessments will be the way of the future, equipping students not just to show that they are proficient in English but also to utilize it successfully in quickly evolving academic, professional, digital, and multilingual contexts.

Author



Dr. Siti Mariam, M.Pd is a lecturer, researcher, and trainer in education with expertise in language education. Her research interests are language assessment, curriculum and materials development, TESOL methodology, inclusive language education, and applied linguistics.

References

- Ajjawi, R., Tai, J., Dollinger, M., Dawson, P., Boud, D., & Bearman, M. (2024). From authentic assessment to authenticity in assessment: Broadening perspectives. *Assessment & Evaluation in Higher Education*, 49(4), 499–510. <https://doi.org/10.1080/02602938.2023.2271193>
- Aryadoust, V. (2023). The vexing problem of validity and the future of second language assessment. *Language Testing*, 40(1), 10–16. <https://doi.org/10.1177/02655322221125204>
- Chuang, P.-L., & Yan, X. (2025). Language assessment in the era of generative artificial intelligence: Opportunities, challenges, and future directions. *System*, 134, 10384. <https://doi.org/10.1016/j.system.2025.103846>
- Fajaria, N. H., Maknun, L., Hasbi, M., Islamiah, N., Male, H., Harida, E. S., Kusmaningrum, M. (2026). *AI in English Language Teaching and Research*. Rizquna.
- Fawns, T., Bearman, M., Dawson, P., Nieminen, J. H., Ashford-Rowe, K., Willey, K., Jensen, L. X., Damşa, C., & Press, N. (2025). Authentic assessment: From panacea to criticality. *Assessment & Evaluation in Higher Education*, 50(3), 396–408. <https://doi.org/10.1080/02602938.2024.2404634>
- Hasbi, M. (2026). Harry Potter and the Last Spell Standing: A gamified pre-exam review activity in the EFL classroom. In *Games in English Language Teaching* (pp. 37-51). Rizquna.
- Hasrol, S. B., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Methods in Applied Linguistics*, 1(3), 100023.

<https://doi.org/10.1016/j.rmal.2022.100023>

- Kofinas, A. K., Tsay, C.-H. H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56, 2522–2549. <https://doi.org/10.1111/bjet.13585>
- Mariam, S., Fadlilah, S., Kepirianto, C. (2025a). Online Assessment of English Competence and Its Washback: Teachers' and Students' Voices. *JSL: International Journal of Social Learning*, 5(3), 574–589. <https://doi.org/10.47134/ijsl.v5i3.437>
- Mariam, S., Fadlilah, S., Kepirianto, C. (2025b). What Are the Washback Effects of Online Students' English Competence Assessment? *AWEJ: Arab World English Journal*, 16(2), 445–458. <https://dx.doi.org/10.24093/awej/vol16n02.25>
- Mariam, S., Kepirianto, C., Fadlilah, S. (2025). Indonesian EFL Teachers' Beliefs and Practices in the Implementation of Authentic Assessment for Speaking Skills. *Forum for Linguistic Studies*, 07(12), 1290–1301. <https://doi.org/10.30564/fls.v7i12.11738>
- Quinlan, K. M., Selleib, G., & Fioruccio, W. (2024). Educationally authentic assessment: Reframing authentic assessment in relation to students' meaningful engagement. *Teaching in Higher Education*, 717–734. <https://doi.org/10.1080/13562517.2024.2394042>
- Zhang, T., Erlam, R., & Botelho de Magalhães, M. (2025). Exploring the dual impact of AI in post-entry language assessment: Potentials and pitfalls. *Annual Review of Applied Linguistics*, 45, 274–293. <https://doi.org/10.1017/S0267190525000030>