

WHEN THE PRODUCT NO LONGER PROVES THE LEARNING: REDESIGNING ASSESSMENT IN ENGLISH LANGUAGE TEACHING UNDER GENERATIVE AI

Bo'riyeva Go'zal To'ychiyevna

Wise school

English language Teacher

E-mail: guzalburiyeva8@gmail.com

<https://doi.org/10.5281/zenodo.22637899>

Qabul qilindi: 1-sentabr 2026-yil • Ma'qullandi: 3-sentabr 2026-yil • Nashr qilindi: 5-sentabr 2026-yil

Abstract.

The paper examines the consequence that generative artificial intelligence carries for assessment in English language teaching. Since a finished essay, translation or scripted dialogue can now be obtained within seconds, the assumption that a completed product testifies to the learner's own competence no longer holds. The paper distinguishes the four language skills according to how far their usual evidence has been weakened, sets out process-based alternatives for each, and proposes a three-band scheme regulating permitted AI use in tasks. It is argued that assessment literacy, rather than detection technology, is the decisive professional response.

Key words:

assessment redesign, generative AI, English language teaching, process evidence, academic integrity, declared use, teacher assessment literacy, feedback.

Introduction

Assessment in language education has long rested on a quiet assumption: that a text submitted by a learner was produced by that learner and therefore shows what the learner can do. Generative models removed this assumption almost overnight. A composition on a set topic, a translation of a passage, a dialogue for a role-play, even a reflective journal entry can now be produced in seconds and in a register the learner has not yet reached [1; 5].

The scale of the change is uneven across the profession. Surveys of university language instructors record limited confidence in using such tools pedagogically and the late arrival of institutional guidance [2], while those preparing future teachers report that their own programmes have not yet been rewritten around the technology [3]. Among pre-service teachers of English, about sixty per cent do not feel able to apply these tools in teaching, and two thirds describe their grasp of the underlying concepts as shallow [6]. Assessment, however, cannot wait for readiness: the work is submitted every week.

Institutional practice has already indicated a direction. An analysis of the guidelines issued by the world's leading universities found that they responded not by prohibition but by shifting emphasis towards process, declared use and academic integrity [4]. UNESCO's guidance on generative AI in education and its competency framework for teachers point the same way, placing human oversight at the centre [7; 8]. The purpose of this paper is to translate that general direction into concrete assessment decisions for the English classroom.

Main part

Which evidence has been weakened. The four skills are not affected equally, and treating them as a single case leads to disproportionate restrictions. Written production is affected most severely, since

its evidence is precisely a detachable artefact. Speaking, by contrast, retains much of its evidential value because performance takes place in real time under interactional pressure. Reading and listening occupy an intermediate position: the comprehension itself cannot be delegated, yet the written answer that reports it can be. Table 1 summarises the distinction together with the process-based evidence that replaces the weakened form.

Table 1

Assessment evidence in the four skills before and after generative AI

Skill	Customary evidence	How far it is weakened	Process-based evidence that replaces it
Writing	the finished text	severely: the whole product can be generated	drafts with tracked revisions; a short oral defence of choices; in-class writing under observation
Speaking	live performance	slightly: real-time interaction cannot be delegated	unrehearsed follow-up questions; interaction with a partner rather than monologue
Reading	written answers about a text	moderately: comprehension is personal, the report of it is not	questions requiring reference to a specific line; oral summary; comparison of two passages
Listening	written answers about a recording	moderately: same asymmetry as reading	note-taking checked in class; retelling; reaction to an unseen segment

Source: compiled by the author on the basis of [1]; [4]; [5].

Regulating use rather than forbidding it. A blanket prohibition is unenforceable and, more importantly, pedagogically wasteful: it denies learners practice in the verification and editing skills they will need regardless. Detection software offers no reliable escape either, since its judgements are probabilistic and its errors fall unevenly on non-native writers. What remains workable is an explicit contract stating for each task which uses are permitted, which must be declared and which are excluded (Table 2).

Table 2

A three-band scheme for permitted AI use in language tasks

Band	What is permitted	Typical task	What the teacher checks
Band 1 — no AI use	no generative tool at any stage	in-class writing, live speaking, controlled test	performance observed directly; conditions guarantee authorship
Band 2 — declared support	brainstorming, checking grammar, explaining a rule; the learner writes the text	home essay, project report, presentation script	declaration of what was used and where; drafts; ability to justify wording
Band 3 — AI as object of study	generation is part of the task itself	editing a generated text; comparing two versions; identifying register and factual errors	quality of the correction and of the reasoning, not of the original output

Source: compiled by the author on the basis of [4]; [7]; [8].

The third band deserves emphasis, since it converts the threat into instructional material. Asking learners to repair a generated text trains precisely the competence that fluent machine output endangers: the ability to notice that a collocation is improbable, that a register sits above the situation, or that a statement is confidently wrong. Research on machine-generated text for second-language writers describes the same double effect — access to fluent English on one side, the flattening of individual voice on the other [5] — and a task built on correction addresses both.

Intervention evidence supports the emphasis on judgement rather than tool handling. In an eight-week programme with 64 teachers of English in which AI work was embedded in genre-based pedagogy, the largest gains fell on critical appraisal, while technical understanding moved only modestly [9]. A twelve-week course for 90 pre-service teachers produced significant gains in critical evaluation and ethical competence alongside technical proficiency, its authors concluding that such instruction belongs inside teacher education rather than outside it [10]. Both results indicate that the assessment response outlined here presupposes prior preparation of the teacher.

Conclusion

1. Generative models have invalidated the finished product as sufficient evidence of learning in written work, while leaving live speaking largely intact; reading and listening stand between these poles, since comprehension is personal but its written report is not.

2. The response should be differentiated by skill rather than uniform: process evidence — drafts, oral defence, observed performance — replaces product evidence where the latter has been weakened.

3. Prohibition and detection software are both unreliable; a workable alternative is an explicit three-band contract distinguishing tasks without AI, tasks with declared support, and tasks in which the generated text is itself the object of study.

4. The third band turns the difficulty into instruction: correcting a generated text develops exactly the evaluative competence that fluent machine output places at risk.

5. None of this is achievable without prepared teachers, and intervention studies show that evaluative judgement — unlike tool handling — grows only through explicit and sustained instruction.

Adabiyotlar, References, Литературы:

1. Kohnke L., Moorhouse B.L., Zou D. ChatGPT for language teaching and learning // RELC Journal. – 2023. – Vol. 54, № 2. – P. 537–550.
2. Kohnke L., Moorhouse B.L., Zou D. Exploring generative artificial intelligence preparedness among university language instructors: A case study // Computers and Education: Artificial Intelligence. – 2023. – Vol. 5. – Art. 100156.
3. Moorhouse B.L., Kohnke L. The effects of generative AI on initial language teacher education: The perceptions of teacher educators // System. – 2024. – Vol. 122. – Art. 103290.
4. Moorhouse B.L., Yeo M.A., Wan Y. Generative AI tools and assessment: Guidelines of the world's top-ranking universities // Computers and Education Open. – 2023. – Vol. 5. – Art. 100151.
5. Warschauer M., Tseng W., Yim S., Webster T., Jacob S., Du Q., Tate T. The affordances and contradictions of AI-generated text for writers of English as a second or foreign language // Journal of Second Language Writing. – 2023. – Vol. 62. – Art. 101071.
6. Karaduman C. Pre-service EFL teachers' perceived AI literacy and competency: The integration of ChatGPT into English language teacher education // SAGE Open. – 2025. – Vol. 15, № 3. – Art. 21582440251379712.
7. UNESCO. Guidance for generative AI in education and research. – Paris: UNESCO, 2023. – 44 p.
8. UNESCO. AI competency framework for teachers. – Paris: UNESCO, 2024. – 66 p.
9. Zhang Y., Samsudin M.A. From familiarity to criticality: Cultivating EFL teachers' AI literacy through an AI-integrated genre-based pedagogy // Education Sciences. – 2026. – Vol. 16, № 1. – Art. 150.
10. Huang T., Wu C., Wu M. Developing pre-service language teachers' GenAI literacy // Discover Artificial Intelligence. – 2025. – Vol. 5, № 1. – Art. 163.
11. Resolution of the President of the Republic of Uzbekistan No. PQ-358 of 14 October 2024 on the Strategy for the Development of Artificial Intelligence Technologies until 2030. – Tashkent, 2024.