

ACADEMIC INTEGRITY AND GENERATIVE AI IN LANGUAGE ASSESSMENT: FORMULATING FAIR EVALUATION STRATEGIES FOR EFL CLASSROOMS

Kumush Obiddin qizi

*English education Student at Kimyo International University in Tashkent
Tashkent, Uzbekistan*

Abstract

Generative artificial intelligence (AI) tools such as ChatGPT can produce fluent English texts in seconds. This creates a serious challenge for language assessment, because teachers can no longer be sure that a student's written work shows the student's own ability. This article discusses academic integrity in English as a Foreign Language (EFL) classrooms in the age of generative AI and proposes fair evaluation strategies. It reviews research on AI and academic integrity, explains why AI-detection software is an unreliable and potentially unfair solution for EFL learners, and presents a practical framework based on clear AI-use rules, process-based assessment, oral verification and reflective tasks. The article argues that fair assessment does not depend on banning AI completely, but on redesigning tasks and being transparent about what is allowed.

Keywords: *generative AI, academic integrity, language assessment, EFL, fairness, assessment design*

1. Introduction

Language teachers have always worried about plagiarism, copying and contract cheating. Generative AI has made this problem larger and more complex. A learner with a low level of English can now submit a well-structured essay with few grammar mistakes, even if the learner could not write it alone. At the same time, AI tools can be a useful support for learning: they can explain grammar, give feedback and offer examples.

For EFL teachers, the situation creates a dilemma. If they ignore AI, assessment results may no longer be valid. If they ban AI strictly and use detection software, they may punish honest students unfairly. Assessment in language classrooms must be valid (it measures the intended skill), reliable (results are consistent) and fair (all learners have equal and reasonable chances to show their ability) (Bachman & Palmer, 1996).

This article has three aims:

1. to explain the main academic integrity risks of generative AI in EFL assessment;
2. to examine the limits of AI-detection tools;

3. to propose a practical framework for fair evaluation.

Research questions

- RQ1: How do EFL teachers and learners understand academic integrity when using generative AI?
- RQ2: Which assessment strategies are seen as most fair and effective in reducing inappropriate AI use?
- RQ3: What conditions make AI-use policies clear and acceptable to learners?

2. Literature Review

2.1 Generative AI and Academic Integrity

Researchers agree that generative AI changes the meaning of academic integrity. Cotton, Cotton and Shipway (2024) discuss the opportunities and risks of ChatGPT in higher education and recommend clear policies, staff and student education, and assessment redesign. Sullivan, Kelly and McLaughlan (2023) also show that universities are still developing their responses, and that rules are often unclear. Perkins (2023) argues that academic integrity policies must be updated to include AI, and that the focus should not be only on punishment.

2.2 Problems with AI-Detection Tools

Many institutions look for technical solutions, such as software that claims to detect AI-written text. However, these tools have serious limits. Liang et al. (2023) found that several popular detectors wrongly classified writing by non-native English speakers as AI-generated more often than writing by native speakers. This finding is especially important for EFL contexts. Learners with simpler vocabulary and more regular sentence patterns may be marked as “suspicious” without any cheating. For this reason, detection results should never be the only evidence in an integrity case.

2.3 Assessment Design and Integrity

Research on cheating suggests that assessment design is a stronger protection than surveillance. Bretag et al. (2019) found that tasks that are authentic, personalized and linked to in-class activity, and that include oral components, are related to lower risk of contract cheating. Dawson (2021) similarly argues that assessment security should combine several methods, not depend on one tool. Black and Wiliam (1998) show that formative assessment with feedback supports learning, which suggests that process-focused tasks are valuable for both integrity and development.

2.4 Levels of AI Use

Perkins, Furze, Roe and MacVaugh (2024) propose an AI Assessment Scale that describes different levels of permitted AI use, from no AI at all to full and creative use of AI. This idea

is helpful for language teachers because it replaces the simple “allowed or banned” view with clear levels for different tasks.

3. Methodology

3.1 Research Design

The study uses a qualitative-dominant mixed-methods design in two stages:

- **Stage 1 (survey):** a questionnaire for EFL teachers and learners on their AI use, understanding of academic integrity and views on fairness.
- **Stage 2 (interviews and document analysis):** semi-structured interviews with teachers, and analysis of existing institutional policies on AI use.

3.2 Participants

The participants are EFL teachers and learners from universities or language centers (for example, 100–150 learners and 10–15 teachers). Participation is voluntary and anonymous. Students are assured that the survey is not used for disciplinary purposes, so that they can answer honestly.

3.3 Data Analysis

Survey data are analyzed with descriptive statistics, and group differences (for example, between proficiency levels) with suitable statistical tests. Interviews and policy documents are analyzed with thematic analysis, focusing on how fairness and integrity are described.

4. A Framework for Fair Evaluation

4.1 Make AI-Use Rules Explicit

Each task should state clearly whether AI is allowed, and how. Following the idea of levels (Perkins et al., 2024), a teacher can use a simple scale:

Level	AI use	Example task
1	No AI	In-class handwritten writing, spoken interview
2	AI for planning only	Brainstorming ideas, then writing alone
3	AI for feedback	Student writes first, then asks AI to explain errors
4	AI as a supported tool	Student edits AI text and explains every change
5	Open and critical use	Student evaluates and improves AI output

Learners should also be required to declare any AI use, for example by adding a short note: which tool, for what purpose, and what they changed.

4.2 Assess the Process, Not Only the Product

Instead of one final essay, teachers can collect outlines, drafts, teacher or peer feedback and a final version. Version history, drafting notes and short reflections show how the learner developed the work. This also supports learning, because feedback during the process is valuable (Black & Wiliam, 1998).

4.3 Add Oral Verification

A short conversation about the written work is a strong and fair way to check authorship. The student explains choices in vocabulary, structure and argument. For EFL classes, this also directly assesses speaking, so it supports the course goals and does not become only a control tool. Oral checks can be used for a sample of students or when there is a reasonable concern.

4.4 Use Personalized and In-Class Tasks

Tasks based on class discussions, local topics, learners’ own experience or data collected by the students are harder for AI to complete in a generic way (Bretag et al., 2019). Some key assessments, such as basic writing and speaking, can be done in class under supervised conditions.

4.5 Do Not Rely on Detectors Alone

Because of the bias risk described by Liang et al. (2023), detector scores should not be used as proof. If a case seems suspicious, the teacher should combine several kinds of evidence (draft history, comparison with earlier work, a conversation with the student) and give the student a fair chance to explain.

4.6 Teach AI Literacy and Integrity

Many students do not know what counts as misuse. Teachers should explain the rules, discuss examples, and teach learners how to use AI responsibly to support, not replace, their own learning. A clear and educational approach is also more fair than punishment without explanation (Cotton et al., 2024).

4.7 Support Equity

Not all learners have the same access to AI tools, internet or digital skills. If AI use is allowed or required, the institution should consider equal access, for example by providing tools in class. Rubrics should focus on the skills being assessed, and should not reward only “perfect” language that AI can easily provide.

5. Results and Discussion

Note: This section is a framework. Real numbers and quotations must come from your own survey and interview data. Please do not publish this section without real results.

5.1 Understanding of integrity (RQ1). Report how participants define misuse, and where teachers’ and learners’ views differ.

5.2 Fairness and effectiveness of strategies (RQ2). Report which strategies (for example, drafts, oral checks, in-class tasks) participants rated as most fair and practical.

5.3 Clarity of policies (RQ3). Report what makes rules understandable, and what causes confusion or anxiety.

5.4 Discussion points. Compare your results with Cotton et al. (2024), Liang et al. (2023) and Bretag et al. (2019), and discuss whether process-based and oral tasks increase workload for teachers.

6. Limitations

Generative AI is changing very quickly, so any specific advice about tools may soon become outdated. Self-reported data on AI use may be incomplete, because students may fear punishment. The proposed strategies also increase teacher workload and may be difficult in large classes. Future research should test these strategies in real classrooms and measure their effect on both integrity and learning.

7. Conclusion

Generative AI cannot be removed from education, and EFL teachers need practical and fair ways to assess language ability. Detection software alone is unreliable and may be especially unfair to non-native writers. A better approach is to combine clear AI-use rules, process-based tasks, oral verification, personalized assignments and AI literacy education. In this way, assessment can protect integrity and, at the same time, help learners develop their own language skills.

References

1. Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.
2. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
3. Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., Saddiqui, S., & van Haeringen, K. (2019). Contract cheating: A survey of Australian university students. *Studies in Higher Education*, 44(11), 1837–1856.
4. Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239.
5. Dawson, P. (2021). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge.
6. Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), 100779.
7. Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2).

8. Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6).

9. Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), 31–40.