

Original Research

Received 25 May 2026

Revised 29 June 2026

Accepted 17 July 2026

Natural Resources for Human Health 2026, Vol. 6 (10s): 614–617

KEYWORDS:

artificial intelligence, assessment, higher education, academic integrity, automated grading, foreign universities.

Traditional and Artificial Intelligence-based Assessment in Foreign Universities

Allahverdiyeva Gunel Zohrab¹, Mammadzada Gulnara Javanshir², Yusubova Maral Aghayar³, Aliyeva Nigar Suleyman⁴

¹PhD, assistant associate professor, Azerbaijan State Pedagogical University
Orcid code: ID 0000-0001-8013-7741, E-mail: allahverdiyevag524@gmail.com

²PhD, senior teacher, Azerbaijan State Pedagogical University
Orcid code: ID 0009 0004 2413 0466, E-mail: gulnare.memmedova.78@mail.ru

³Senior teacher, Azerbaijan State Pedagogical University
Orcid code: ID 0009-0004-2930-112X, E-mail: maral060687@gmail.com

⁴Teacher, Azerbaijan State Pedagogical University
Orcid code: ID 0009-0005-0103-1081, E-mail: nigar.aliyeva97.97@mail.ru

ABSTRACT: The rapid advancement of artificial intelligence (AI) is fundamentally reshaping how student learning is assessed in higher education institutions worldwide. This article provides a comparative analysis of traditional assessment methods (written examinations, essays, oral defenses, project-based evaluation) and AI-based assessment tools (automated grading systems, large language model feedback, AI-proctoring) as applied in foreign universities. Drawing on recent empirical studies, systematic reviews, and institutional practices from universities such as Cornell, Purdue, UC San Diego, and Gothenburg, the study identifies the strengths, limitations, and academic-integrity implications of both approaches. The findings suggest that while AI-based assessment offers clear advantages in speed, scalability, and immediacy of feedback, human evaluation retains an edge in assessing complex, creative, and context-dependent work. The article concludes that a hybrid model — combining AI-assisted tools with human oversight — represents the most promising direction for future assessment practice.

1. INTRODUCTION

Assessment has historically served as the principal means of measuring student knowledge and skills in higher education, taking shape over the years through written examinations, essays, tests, coursework, and oral defenses. However, the rapid spread of generative artificial intelligence systems such as ChatGPT, released to the public in late 2022, has fundamentally shaken the traditional assessment paradigm: the possibility that students could complete assignments with the help of AI has raised serious concerns regarding academic integrity.

At the same time, artificial intelligence is not merely a threat but also acts as a tool that is transforming the assessment process itself. Foreign universities — particularly higher education institutions in the United States, the United Kingdom, Australia, and the Scandinavian countries — have begun actively implementing Gradescope, AI-based essay-grading systems, automated code-checking tools, and large language model (LLM)-based formative feedback instruments. The purpose of this article is to systematically compare traditional and AI-based assessment approaches, identify the strengths and weaknesses of each, and discuss future directions for development based on the actual practice of foreign universities.

2. LITERATURE REVIEW

Over the past two to three years, the number of studies on the role of artificial intelligence in educational assessment has grown sharply. Systematic reviews show that AI-based assessments demonstrate a strong correlation with the scores given by human experts, particularly in measuring cognitive indicators such as inference, originality, and clarity of explanation (IIETA, 2024). At the same time, researchers warn that excessive reliance on AI may weaken students' independent problem-solving skills.

A study published in *Frontiers in Education* shows that traditional assessment tools — standardized examinations and essays — are becoming increasingly vulnerable to AI-assisted plagiarism, which is prompting instructors to shift toward "AI-resistant" assessment formats such as podcasts, multimedia presentations, and real-world problem-solving tasks (*Frontiers in Education*, 2024).

A comparative study published by Taylor & Francis compared grades given by AI chatbots, instructors, and peers on the same assignments and found that artificial intelligence systems tend to award higher grades than human evaluators, whereas instructor and peer grades are more closely aligned with one another. The authors emphasize that a balanced approach combining the effectiveness of AI with the contextual sensitivity of human judgment is more appropriate (*Assessment & Evaluation in Higher Education*, 2025).

A systematic review published in the journal *AI and Ethics* (Springer) highlights the need to reconsider assessment strategies in the context of academic integrity: the authors note the growing popularity of competency-based and project-based forms of "authentic assessment" that prioritize critical thinking, problem-solving, and creativity over rote memorization (*AI and Ethics*, Springer, 2025).

Finally, an empirical study conducted at the University of Gothenburg found that ChatGPT's ability to grade master's-level exam papers produced results comparable to those of a single human grader, although the researchers currently recommend it as a supportive tool rather than an independent grading method (Flodén, *British Educational Research Journal*, 2025).

3. RESEARCH METHODOLOGY

The study is based on a qualitative analysis method and employs a comparative-analytical approach applied to scholarly articles, systematic reviews, conference proceedings, and official university reports published between 2023 and 2025. Sources were selected according to the following criteria: (1) direct relevance to the application of AI-based assessment tools; (2) conduct within the context of higher education; (3) publication in peer-reviewed journals or reputable scholarly repositories (journals indexed in Scopus/Web of Science, arXiv). Based on the collected data, traditional and AI-based assessment methods were systematically compared across eight key criteria (see Table 1).

4. TRADITIONAL ASSESSMENT METHODS

Traditional assessment in foreign universities primarily encompasses the following forms: written and oral examinations, essays and analytical writing assignments, coursework and thesis papers, laboratory reports, oral presentations and defenses, and practical (clinical, engineering) examinations. The main advantage of these methods lies in the instructor's ability to exercise contextual judgment, take into account the student's individual development, and evaluate complex, multi-layered argumentation. At the same time, traditional assessment is labor-intensive, characterized by delayed feedback, and may be affected by the instructor's subjective factors (fatigue, mood, bias).

5. ARTIFICIAL INTELLIGENCE-BASED ASSESSMENT METHODS

AI-based assessment systems fall into several main categories: automated test- and code-grading platforms (e.g., Gradescope), large-language-model-based essay and project assessment, AI-detection and proctoring systems that monitor academic integrity, and personalized, adaptive formative feedback tools. The Gradescope platform is widely implemented at universities such as Cornell, Purdue, UC San Diego, Florida, Rutgers, and Indiana University, offering features such as automatic answer grouping and automated grading of certain question types (Ohio State University, ASC Office of Distance Education, 2025). The main advantages of these systems are scalability, speed, and consistency — thousands of papers can be graded in a short time using the same criteria.

However, studies show that AI-based assessment systems have their own limitations: a tendency to grade weak papers more leniently and strong papers more harshly than warranted, the risk of bias stemming from training data, and, at times, the risk of "hallucinating" non-existent references or facts (Ohio State University, 2025). For this reason, researchers recommend using AI not yet as an independent assessment tool but as a supportive instrument under human oversight.

6. COMPARATIVE ANALYSIS: TABLE 1

Table 1. Comparison of traditional and AI-based assessment across key criteria

Criterion	Traditional Assessment	AI-Based Assessment
Speed and scalability	Low; requires manual instructor effort	High; thousands of papers can be graded in parallel
Feedback turnaround	Usually delayed by days or weeks	Instant or within minutes
Risk of subjectivity	Affected by the instructor's mood and fatigue	Risk of algorithmic (data-driven) bias
Assessing creativity and complex ideas	Strong; contextual judgment possible	Limited; difficult to fully evaluate nuanced, original ideas
Academic integrity monitoring	Limited to traditional plagiarism checks	Partially effective at detecting AI-generated text, but can be unreliable
Cost and resource requirements	High long-term labor costs	High initial investment, cost-efficient afterward
Transparency	Criteria usually clear, but documentation may vary	"Black box" effect; decision logic is less explained

7. PRACTICES IN FOREIGN UNIVERSITIES

US universities such as Cornell, Purdue, UC San Diego, Florida, Rutgers, and Indiana University use hybrid assessment systems via the Gradescope platform that accommodate both traditional (paper-based, scanned) and online exam formats (ASC Office of Distance Education, 2025). In an empirical study conducted at the University of Gothenburg in Sweden, exam papers from a master's-level logistics course were graded using ChatGPT, and the results were comparable to those of a human grader (Flodén, 2025).

In another experiment conducted in the field of environmental data science, a hybrid assessment model called "FACT" (Fundamental, Applied, Conceptual, critical Thinking) was implemented: AI tools were used to assess applied project work and critical-thinking components, while core concepts were verified through traditional methods (Frontiers in Education, 2025). This model shows that foreign universities increasingly favor a selective integration strategy rather than full replacement — AI is retained for certain types of skills, while traditional methods are preserved for others.

In the area of academic integrity protection, universities employ both AI-detection tools and "AI-resistant" task design (oral defenses, process-oriented assessment, reflective writing tied to feedback). Nevertheless, researchers emphasize that the detection of AI-generated text remains unreliable and cannot serve as a standalone solution (Frontiers in Education, 2024; Springer AI and Ethics, 2025).

8. DISCUSSION

The analysis shows that a relationship of functional complementarity, rather than competition, is forming between AI-based and traditional assessment. The main strength of AI lies in evaluating scalable, standardized tasks with objective answers (tests, code, mathematical problems). Traditional (human) assessment, in turn, retains its advantage for tasks requiring complex argumentation, creativity, ethical judgment, and consideration of individual context.

The experience of foreign universities shows that a successful implementation strategy converges on a "hybrid model": AI is used for preliminary assessment, large-scale feedback, and reducing administrative burden, while the final decision remains under human oversight (the "human-in-the-loop" principle). This approach both increases efficiency and limits the impact of algorithmic bias and transparency problems.

9. CONCLUSION

The integration of artificial intelligence into assessment practice in higher education has become an irreversible process, and the experience of foreign universities clearly demonstrates this. The findings of this study reveal that traditional assessment is not being entirely replaced but rather complemented by AI-based tools. The main directions for future research are as follows: developing methods to reduce algorithmic bias in AI-based assessment systems, conducting long-term empirical verification of the effectiveness of hybrid assessment models across various disciplines, and investigating the psychological and pedagogical aspects of students' trust in AI-based assessment. The recommendation for higher education institutions is to develop balanced assessment frameworks that preserve human oversight, ensure transparency, and safeguard academic integrity.

REFERENCES

- [1] AI and Auto-Grading in Higher Education: Capabilities, Ethics, and the Evolving Role of Educators. (2025). ASC Office of Distance Education, The Ohio State University. <https://ascode.osu.edu/news/ai-and-auto-grading-higher-education-capabilities-ethics-and-evolving-role-educators>
- [2] AI-Driven Assessment Systems in Higher Education: Effectiveness for Enhancing Critical Thinking and Creativity. (2024). *Ingénierie des Systèmes d'Information (ISI)*, 30(6). <https://www.ieta.org/journals/isi/paper/10.18280/isi.300624>
- [3] AI-resistant assessments in higher education: practical insights from faculty training workshops. (2024). *Frontiers in Education*. <https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2024.1499495/full>
- [4] Balancing AI-assisted learning and traditional assessment: the FACT assessment in environmental data science education. (2025). *Frontiers in Education*. <https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1596462/full>
- [5] Flodén, J. (2025). Grading exams using large language models: A comparison between human and AI grading of exams in higher education using ChatGPT. *British Educational Research Journal*. <https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/berj.4069>
- [6] Generative AI vs. instructor vs. peer assessments: a comparison of grading and feedback in higher education. (2025). *Assessment & Evaluation in Higher Education*. <https://www.tandfonline.com/doi/full/10.1080/02602938.2025.2487495>
- [7] Kani, R. M., Maharram, V. K., Dhar, S., Rajendran, P., & Ahmed, S. A. (2025, May). Automating Grading to Enhance Student Feedback and Efficiency in Higher Education with a Hybrid Ensemble Learning Model. In *2025 Global Conference in Emerging Technology (GINOTECH)* (pp. 1-6). IEEE. **Doi.** [10.1109/GINOTECH63460.2025.11076638](https://doi.org/10.1109/GINOTECH63460.2025.11076638)
- [8] Karimli, V. M., Khudaverdiyeva, T. S., Huseynova, F., Zahid, F., Bayramova, N. A., Honfi, V. S., & Narayana, V. S. K. (2025). The role of mobile computing in adaptive testing for English language learners: Personalizing assessment to improve outcomes. In *Forum for Linguistic Studies* (Vol. 7, No. 6, pp. 149-160). <https://doi.org/10.30564/fls.v7i6.9663>
- [9] Machine vs Machine: Using AI to Tackle Generative AI Threats in Assessment. (2025). arXiv preprint. <https://arxiv.org/pdf/2506.02046>
- [10] Newman, N. (2024). Plagiarism, proctoring and co-pilots: university assessment in the AI era. Medium. <https://medium.com/@naxn/plagiarism-proctoring-and-co-pilots-university-assessment-in-the-ai-era-70ae1a469c97>
- [11] P. Keshwani, V. Kiran Kumar Ravi, M. Kanmani, V. Karimli Maharram, S. Kumari and M. Shunmugasundaram, "Hybrid Deep CNN-ELM Based Auto-Grading System for Reducing Educator Workload and Enhancing Student Performance in Higher Education," *2025 3rd World Conference on Communication & Computing (WCONF)*, Raipur, India, 2025, pp. 1-6, doi: [10.1109/WCONF64849.2025.11233444](https://doi.org/10.1109/WCONF64849.2025.11233444)
- [12] Redefining student assessment in AI-infused learning environments: a systematic review of challenges and strategies for academic integrity. (2025). *AI and Ethics*. <https://link.springer.com/article/10.1007/s43681-025-00871-w>
- [13] Tian, Z., Liu, A., Esbenshade, L., & Sarkar, S. (2025). Implementation Considerations for Automated AI Grading of Student Work. arXiv preprint. <https://arxiv.org/pdf/2506.07955>