Registered in Germany Europa Revue Journal
ISSN: 1862-6564 Vol. 6 No. 10 (2026) Published October 10, 2026

THE USE OF ARTIFICIAL INTELLIGENCE IN DEVELOPING RUBRICS FOR STUDENT ASSESSMENT IN HIGHER EDUCATION

Zayniyeva Gulchaman Ilhomjon qizi
Teacher, National university of Uzbekistan named after Mirzo Ulugbek
Department of foreign language and literature
gulchamanzayniyeva@gmail.com
+99897 002 53 71

Abstract

The rapid diffusion of generative artificial intelligence (GenAI) is reshaping assessment practices in higher education and is creating new opportunities for the design of transparent, scalable, and learning-oriented assessment instruments. This conceptual research article examines how artificial intelligence can support the development of analytic rubrics while preserving teacher agency and assessment validity. The study uses an integrative literature review and design-oriented conceptual analysis of research on scoring rubrics, AI-supported assessment, generative AI, and international guidance for responsible AI in education. Rather than treating AI as an autonomous grader, the article conceptualizes AI as a rubric-design assistant operating within a human-in-the-loop model. The proposed framework comprises seven stages: learning-outcome analysis, task analysis, AI-assisted criterion generation, descriptor generation, expert validation, pilot calibration, and iterative refinement.

The analysis identifies four principal opportunities: reducing rubric-development workload, improving criterion coverage, facilitating differentiated descriptors, and supporting alignment between learning outcomes and assessment tasks. At the same time, the study identifies significant risks, including generic descriptors, construct underrepresentation, bias, over-reliance on AI-generated language, and inappropriate transfer from rubric generation to autonomous grading. Recent evidence suggests that supervised AI-rubric systems can achieve useful levels of consistency in particular contexts, while broader evidence on AI-based grading remains mixed and context dependent. The article therefore argues that the educational value of AI lies not in

replacing professional assessment judgment but in augmenting teachers' capacity to design, test, and refine assessment instruments. The proposed framework provides a practical and theoretically grounded basis for responsible AI-assisted rubric development in higher education.

Keywords

artificial intelligence; generative AI; assessment rubrics; higher education; AI-assisted assessment; formative assessment; assessment validity; teacher agency; ChatGPT; learning outcomes.

Introduction

Assessment in higher education is undergoing a fundamental transformation. Universities increasingly need assessment systems capable of measuring not only factual knowledge but also higher-order competencies such as critical thinking, communication, creativity, problem solving, collaboration, and disciplinary reasoning. At the same time, the rapid adoption of generative artificial intelligence (GenAI) has challenged conventional assumptions about written assignments, feedback, authorship, and grading. Recent research frames the current period as a need to redesign assessment rather than simply attempt to detect AI-generated text (Ahangama, 2026; Kofinas et al., 2025).

Within this changing environment, rubrics have renewed importance. A well-designed rubric makes the construct being assessed visible by specifying criteria, levels of performance, and descriptors. Rubrics can support scoring consistency, feedback, self-regulated learning, and transparency when their design and implementation are appropriate (Andrade, 2000; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). However, rubric development is itself a complex professional activity. Teachers must translate learning outcomes into observable performance dimensions, distinguish meaningful levels of achievement, avoid construct-irrelevant criteria, determine appropriate weighting, and ensure that the language is understandable to students.

Generative AI introduces a potentially powerful design aid. Large language models can rapidly propose criteria, descriptors, alternative formulations, weighting schemes, examples, and alignment checks. Recent studies have begun to investigate AI-generated and AI-assisted rubrics directly. For example, supervised generative-AI rubric models

have been reported to support consistent and pedagogically aligned assessment in particular postgraduate contexts, while newer experimental evidence suggests that structured AI-generated rubrics can sometimes achieve scoring comparability with human-created rubrics under controlled conditions. Nevertheless, these findings do not establish universal validity across disciplines, tasks, student populations, or AI systems. The central problem, therefore, is not whether AI can produce a rubric. It clearly can. The more consequential question is whether AI-assisted rubric development can produce instruments that are valid, transparent, context-sensitive, and educationally useful when subjected to expert scrutiny. This distinction is important because an apparently sophisticated rubric can still measure the wrong construct or create artificial distinctions between performance levels.

UNESCO's AI Competency Framework for Teachers emphasizes human-centred AI use, ethics, AI foundations and applications, AI pedagogy, and professional learning. These principles support an approach in which teachers remain accountable for educational decisions while AI provides assistive capabilities. The present article applies this human-centred perspective specifically to rubric development.

The purpose of this article is to develop a theoretically grounded and practically applicable framework for AI-assisted rubric development in higher education. Four research questions guide the analysis:

RQ1. What functions can generative AI perform in the development of higher-education assessment rubrics?

RQ2. What pedagogical and operational benefits can AI-assisted rubric development provide?

RQ3. What validity, reliability, ethical, and contextual risks may arise?

RQ4. What human-in-the-loop framework can support responsible and effective AI-assisted rubric development?

Literature Review

Rubrics as Assessment and Learning Instruments

Rubrics are commonly understood as structured scoring guides that articulate assessment criteria and levels of performance. Their value extends beyond assigning marks. When

criteria are shared with students before or during an assessment task, rubrics can clarify expectations and support planning, self-assessment, peer assessment, and revision. This formative potential is particularly relevant to higher education because students increasingly need to regulate their own learning.

The literature also demonstrates that rubrics are not automatically valid or reliable simply because they contain multiple criteria. Jonsson and Svingby (2007) found that scoring rubrics can improve reliability, particularly when criteria are specific and raters are appropriately trained, but effects vary according to rubric design and implementation. Panadero and Jonsson (2013) similarly concluded that rubrics can support formative assessment and self-regulated learning, while emphasizing the importance of how rubrics are introduced and used.

Validity remains a central concern. Humphry and Heldsinger (2014) demonstrated that common structural features of rubrics may create threats to validity. This finding has direct implications for AI-generated rubrics: the ability to generate a neat table of criteria should not be mistaken for evidence that the rubric represents the intended construct.

Generative AI and Assessment

Generative AI systems can produce, transform, classify, summarize, and evaluate natural-language content. These capabilities create opportunities for assessment design, feedback, and learning analytics, but also introduce challenges relating to academic integrity, transparency, privacy, bias, and human oversight. Recent work in higher education increasingly emphasizes assessment redesign and AI-aware learning design rather than a narrow focus on detecting AI use.

A key distinction is therefore necessary between AI used for assessment design and AI used for assessment judgment. The former includes brainstorming criteria, generating descriptor drafts, simplifying language, and checking alignment. The latter involves assigning judgments or grades to individual students. The second application carries higher stakes and requires stronger evidence of validity, reliability, fairness, and accountability.

Research Gap

Although the literature on AI in education has expanded rapidly, three gaps remain relevant to rubric development. First, much of the discussion addresses AI-based grading or academic integrity rather than the upstream process of constructing the assessment instrument itself. Second, studies that evaluate AI-generated rubrics often examine specific disciplines or controlled tasks, limiting generalization. Third, there is a need for a practical framework that explicitly separates AI's generative contribution from the teacher's validation responsibility.

The present article addresses this gap by conceptualizing rubric development as a collaborative design process rather than a fully automated assessment pipeline.

Methodology

This study adopts a conceptual, integrative literature-review design. It synthesizes established scholarship on scoring rubrics with recent research on generative AI and higher-education assessment, complemented by international guidance on responsible AI in education. The analysis is organized around four dimensions: (1) rubric design and assessment validity; (2) AI capabilities relevant to rubric construction; (3) empirical evidence on AI-assisted assessment and rubric use; and (4) ethical and pedagogical requirements for human oversight.

The purpose of the methodology is not to claim an experimental effect size. No fictitious student sample, intervention group, or statistical outcome is introduced. Instead, the article develops a design framework grounded in the converging requirements identified across the literature. This approach is appropriate for establishing a conceptual model that can subsequently be tested through empirical research.

For future empirical validation, the proposed framework can be evaluated through expert-rating studies, inter-rater reliability analysis, rubric-comparison experiments, student perception surveys, and controlled comparisons between teacher-created and AI-assisted rubrics.

AI-Assisted Rubric Development Framework

The proposed Human-in-the-Loop AI Rubric Development Framework consists of seven sequential but iterative stages.

Stage	Teacher responsibility	Potential AI contribution	Quality-control question
1. Learning-outcome analysis	Define intended knowledge, skills and competencies.	Suggest measurable dimensions and identify possible ambiguity.	Does each criterion represent an intended learning outcome?
2. Assessment-task analysis	Specify task demands, genre, discipline and student level.	Extract task requirements and propose assessment dimensions.	Does the rubric assess the actual task rather than a generic task?
3. Criterion generation	Select and prioritize criteria.	Generate candidate criteria and alternative structures.	Are criteria relevant, non-overlapping and construct-valid?
4. Descriptor generation	Define meaningful performance distinctions.	Draft level descriptors and observable indicators.	Are descriptors specific, observable and appropriately differentiated?
5. Expert validation	Review validity, fairness, language, weighting and context.	Identify inconsistencies, ambiguities and possible omissions.	Would qualified teachers interpret the rubric similarly?
6. Pilot calibration	Test the rubric on representative student work.	Compare descriptor application across examples and flag difficult cases.	Does the rubric produce defensible and consistent judgments?

7. Iterative refinement	Approve the final version and monitor outcomes.	Suggest revisions based on identified weaknesses.	Does the revised rubric improve clarity without narrowing the construct?

Designing High-Quality AI-Assisted Rubrics

Criterion Design

Criteria should represent meaningful dimensions of the target construct. For an argumentative academic essay, argument quality and evidence use may be central constructs, whereas formatting may be secondary. AI can generate a broad list of possible criteria, but teachers must prevent criterion inflation. Too many criteria increase cognitive and scoring load and may encourage fragmented rather than holistic judgments.

Descriptor Design

High-quality descriptors should be observable, discriminating, and sufficiently specific to support consistent interpretation. Descriptors such as 'excellent understanding' or 'good language' are weak because they do not explain what evidence would justify the judgment. AI prompts should therefore explicitly request observable indicators, avoidance of vague evaluative adjectives, and clear distinctions between performance levels.

Alignment and Weighting

AI can assist with constructive alignment by mapping learning outcomes to task requirements and rubric criteria. However, weighting should reflect the educational importance of each outcome rather than the amount of text devoted to it. Teachers should verify that the rubric does not unintentionally reward superficial features, such as grammatical accuracy, more heavily than the substantive construct being taught.

Student Accessibility

Rubrics should be intelligible to students. AI can be useful for producing student-friendly versions without changing the underlying assessment standards. Providing the

rubric before assessment can make expectations transparent and can support self-regulated learning.

Example: AI-Assisted Rubric for University Academic Writing

The following example demonstrates how the proposed framework can be operationalized for a university-level argumentative essay. It is presented as a design example rather than as an experimentally validated instrument.

Criterion	**Weight**	**Excellent (4)**	**Good (3)**	**Developing (2)**	**Beginning (1)**
Argument & critical thinking	25%	Presents a focused, logically developed argument and evaluates competing perspectives.	Presents a clear argument with relevant analysis and some evaluation.	Argument is identifiable but analysis is uneven or mainly descriptive.	Argument is unclear, unsupported, or predominantly descriptive.
Organization & coherence	20%	Ideas progress logically; paragraphs are purposeful and cohesive.	Overall organization is logical with minor lapses in cohesion.	Organization is generally understandable but connections are sometimes weak.	Ideas are poorly sequenced and relationships between points are unclear.
Evidence & support	20%	Uses relevant, credible evidence	Uses relevant evidence with generally	Evidence is limited, unevenly integrated, or	Evidence is absent, inappropriate, or poorly

		and integrates it analytically.	effective integration.	insufficiently evaluated.	connected to claims.
Academic vocabulary	15%	Uses precise, varied, and context-appropriate academic language.	Uses appropriate academic vocabulary with occasional repetition.	Vocabulary is adequate but limited or sometimes imprecise.	Limited vocabulary frequently restricts precise expression.
Grammar & syntax	15%	Uses accurate and varied structures with errors that are rare and non-disruptive.	Generally accurate structures; minor errors do not affect meaning.	Frequent errors occur, but meaning is usually recoverable.	Persistent errors substantially interfere with clarity.
Academic conventions	5%	Citation, formatting and academic conventions are consistently applied.	Minor inconsistencies in conventions.	Several inconsistencies are present.	Major or repeated problems with academic conventions.

Benefits of AI-Assisted Rubric Development:

1. **Efficiency:** AI can reduce the time required to generate initial rubric drafts, alternative formulations, and differentiated descriptors. This is particularly useful for teachers who design several assessments across different courses.
2. **Design Diversity:** AI can propose analytic, holistic, task-specific, or competency-oriented structures, allowing teachers to compare alternatives before selecting a final design.
3. **Differentiation:** descriptors can be adapted for different proficiency levels, disciplines, genres, and learning contexts.
4. **Alignment Support:** AI can compare learning outcomes, task instructions and rubric criteria, helping teachers identify possible gaps or redundancies.
5. **Accessibility and Communication:** AI can convert technical rubric language into student-friendly language while preserving the intended standard, provided the teacher verifies that meaning has not been weakened.

Risks and Limitations:

1. **Genericity and Surface Quality:** A fluent AI-generated rubric may appear authoritative while containing vague or redundant criteria. Linguistic polish is not evidence of assessment validity.
2. **Construct Underrepresentation:** AI may privilege easily describable features such as grammar, structure, or vocabulary while underrepresenting complex constructs such as originality, disciplinary reasoning, or critical evaluation.
3. **Bias and Fairness:** AI-generated criteria may reflect assumptions embedded in training data. Teachers must examine whether criteria disadvantage particular linguistic, cultural, or disciplinary practices 1unrelated to the intended construct.
4. **Automation Bias:** teachers may accept AI-generated recommendations simply because they are presented confidently. Human review must therefore be substantive rather than ceremonial.
5. **Privacy and Governance: s**tudent work and personal information should be handled according to institutional policy and applicable data-protection requirements. AI use should be transparent where disclosure is required.

6. **Transfer from Rubric Design to Automated Grading:** evidence that AI can help generate a rubric does not establish that AI can reliably assign final grades. These are distinct claims requiring separate validation.

Discussion

The findings of the conceptual analysis support a qualified rather than technological-deterministic view of AI-assisted assessment. Generative AI has genuine value in the design phase because rubric development contains several language-intensive and iterative tasks that large language models can perform efficiently. Yet the core assessment problem is not text generation; it is construct representation and defensible judgment.

The proposed framework therefore treats AI as a design partner operating under professional supervision. This position is consistent with UNESCO's human-centred approach, which emphasizes human agency, ethics, AI pedagogy, and professional learning as core teacher competencies. It also reflects recent empirical work showing that supervised AI-rubric systems can perform promisingly in specific educational contexts while broader AI-based grading evidence remains context dependent.

A particularly important implication is that rubric quality should be evaluated at two levels. The first is internal design quality: clarity, relevance, coverage, level differentiation, and alignment. The second is use quality: whether teachers can apply the rubric consistently and whether students can understand and use it to improve performance. Future empirical studies should therefore move beyond judging whether an AI-generated rubric 'looks good' and instead test whether it improves inter-rater agreement, student understanding, feedback usefulness, and assessment validity.

The framework also suggests a shift in teacher professional development. AI literacy should not be reduced to prompt-writing skills. Teachers need the capacity to critique AI outputs, recognize construct-irrelevant features, evaluate evidence of reliability and validity, protect student data, and make accountable final decisions. UNESCO’s 2024 AI Competency Framework for Teachers explicitly identifies these broader dimensions of professional competence.

Practical Recommendations:

Begin every rubric-design process with clearly stated learning outcomes rather than with an AI prompt.

Provide the AI system with sufficient context: course level, discipline, task, student proficiency, learning outcomes, criteria requirements, and assessment purpose.

Ask for observable descriptors and explicit distinctions between performance levels.

Use AI to generate alternatives, not to determine the final rubric automatically.

Review criteria for construct validity, fairness, redundancy, and appropriate weighting.

Pilot the rubric on representative student work before high-stakes use.

Document major AI-assisted design decisions when institutional policy requires transparency.

Do not transfer evidence about AI-assisted rubric generation to claims about autonomous grading.

Protect student privacy and avoid unnecessary inclusion of identifiable information in AI prompts.

Develop teacher AI literacy around assessment validity, bias, ethics, and human oversight.

Limitations of the Present Study and Directions for Future Research

The principal limitation of this article is its conceptual design. It does not provide an experimental estimate of the effect of AI-assisted rubric development on grading reliability, teacher workload, or student achievement. The framework should therefore be regarded as a theoretically grounded design proposition rather than a demonstrated causal intervention.

Future studies should test the framework empirically across disciplines and proficiency levels. A particularly useful design would compare teacher-created, AI-assisted, and fully AI-generated rubrics using the same student work. Outcomes could include inter-rater reliability, agreement coefficients, rubric usability, teacher time-on-task, student perceptions, feedback quality, and criterion-level validity. Mixed-method studies could combine quantitative scoring evidence with teacher interviews and student focus groups.

Longitudinal research could additionally examine whether repeated use of transparent AI-assisted rubrics influences students' self-assessment and self-regulated learning.

Conclusion

Artificial intelligence offers meaningful opportunities for improving the efficiency and flexibility of rubric development in higher education. Its strongest contribution lies in generating and revising candidate criteria, descriptors, alternative rubric structures, and alignment checks. However, the production of fluent assessment language should not be confused with the production of valid assessment instruments.

This article proposes a Human-in-the-Loop AI Rubric Development Framework in which teachers retain responsibility for construct definition, validation, contextual adaptation, pilot calibration, and final approval. The framework addresses the central tension between technological efficiency and assessment quality by assigning complementary roles to AI and educators.

The broader implication is that responsible AI integration in higher education should strengthen, rather than diminish, professional assessment expertise. AI can accelerate design work, but validity, fairness, contextual sensitivity, and accountability remain fundamentally human responsibilities. Accordingly, AI-assisted rubric development should be implemented as supervised educational design, supported by institutional governance, teacher AI literacy, empirical validation, and continuous review.

References

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. Educational Leadership, 57(5), 13–18.

Humphry, S. M., & Heldsinger, S. A. (2014). Common structural design features of rubrics may represent a threat to validity. Educational Researcher, 43(5), 253–263. https://doi.org/10.3102/0013189X14542154

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. Educational Research Review, 2(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

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(6), 2522–2549. https://doi.org/10.1111/bjet.13585

Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. Educational Research Review, 9, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. Educational Research Review, 30, 100329. https://doi.org/10.1016/j.edurev.2020.100329

Brufau Alvira, N., Bannister, P., & Santamaría Urbieta, A. (2025). Validating the PANDORA GenAI Susceptibility Rubric for Higher Education Assessment: A Field Test of All Translation and Interpreting BA Assignments. Higher Education Quarterly, 79(4), e70056. https://doi.org/10.1111/hequ.70056

Ahangama, N. (2026). Designing assessments in the generative AI era: A tailored assessment framework for ICT tertiary education. International Journal of Educational Technology in Higher Education, 23, Article 9. https://doi.org/10.1186/s41239-026-00582-0

Gupta, V., & Nyamapfene, A. (2025). Integrating Generative AI in Academic Assessments: Practical Recommendations on AI Usage and Enhancing Knowledge Sharing. Internet Reference Services Quarterly, 29(2), 213–223. https://doi.org/10.1080/10875301.2025.2456903

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.

UNESCO. (2024). AI competency framework for teachers. UNESCO.