ELEVATE: JOURNAL OF EDUCATION AND LEARNING
ISSN: XXXX-XXXX | Volume 1 No 1 2026

Authentic Assessment in AI-Infused Learning Environments: An Evidence-Centered Design Framework and Rubric Toolkit for Academic Integrity

Arben Hoxha[1], Elira Leka[2]
Department of Education Sciences, University of Tirana, Tirana, Albania
Department of Information Technology, Aleksandër Moisiu University of Durrës, Durrës, Albania
Corresponding Author Email: arben.hoxha@unitir.edu.al

Abstract
Generative AI tools have destabilized traditional take-home assessment by lowering the cost of producing fluent text, code, and problem solutions. Institutional responses often oscillate between prohibition and permissive use, yet both approaches fail when assessment design does not specify what counts as credible evidence of learning. This article proposes a practical framework for assessment integrity in AI-infused learning environments that shifts attention from detection to design. Using an integrative synthesis of research on authentic assessment, constructive alignment, academic integrity, and emerging guidance on generative AI, we develop an evidence-centered assessment design workflow and a rubric toolkit that make acceptable AI use transparent while preserving the core purpose of assessment: eliciting student thinking. The framework operationalizes five design decisions: defining outcome-relevant evidence, setting AI-use boundary conditions, embedding process traces and checkpoints, using rubric criteria that reward disclosure and reasoning, and adding verification moments such as oral defense or short in-class microtasks. We present a model (Figure 1) and a rubric matrix (Table 1) that can be adapted across disciplines for essays, projects, laboratory reports, and portfolios. The contribution is an implementation-ready package that reduces incentives for misuse, supports equity through clear rules and scaffolding, and enables program-level quality assurance through calibration. We conclude with implications for policy, staff development, and future research on learning outcomes in hybrid human–AI work practices.

Keyword
Authentic Assessment; Generative AI; Academic Integrity; Evidence-Centered Design; Constructive Alignment.

1. Introduction
Generative AI has accelerated a long-running assessment dilemma: many high-stakes tasks reward polished output more than learning processes. When a model can draft essays, generate code, and summarize readings in seconds, output-only evidence becomes increasingly ambiguous. The resulting anxiety is visible in rapid policy shifts, increased reliance on surveillance tools, and renewed calls for invigilated examinations.

Received: 15 January 2026 **Revised:** 18 February 2026 **Published:** 03 March 2026

© Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited.

Yet the core problem is not new technology alone; it is the mismatch between intended learning outcomes and the forms of evidence that assessments collect (Biggs, 1996; Wiggins & McTighe, 2005). If assessment evidence does not require students to externalize reasoning, make disciplinary judgments, and demonstrate ownership of decisions, then academic integrity becomes difficult to sustain, regardless of whether AI is permitted or banned. A design response is therefore needed that treats integrity as a property of assessment systems rather than a trait of individuals.

Research on academic integrity emphasizes that misconduct is shaped by opportunity structures, assessment conditions, and institutional culture, not only by student morality (Bretag, 2016; Newton, 2016). Work on contract cheating and plagiarism shows that risk increases when tasks are generic, under-scaffolded, and assessed primarily through final products (Newton, 2018; Bretag et al., 2019). Conversely, integrity improves when tasks are contextualized, when students receive formative feedback, and when assessments require personal judgment and iterative development (Sambell et al., 2013; Carless, 2015). These findings align with authentic assessment traditions, which argue that learning is best evidenced through meaningful performance tasks embedded in disciplinary practices (Wiggins, 1990). In AI-infused contexts, authentic assessment must also clarify how AI may or may not be used as a tool within those practices.

Many institutions currently frame generative AI as a detection problem. However, detection tools face technical limits, can produce false positives, and may create inequities for multilingual students and those using accessibility tools (Eaton, 2023). A detection-first approach can also undermine trust and shift attention away from learning. An alternative is to redesign assessments so that the most valued evidence is difficult to outsource and easy to verify through low-burden mechanisms. This aligns with constructive alignment, which positions assessment as the primary driver of student learning behavior (Biggs & Tang, 2011). If assessments reward transparency, reasoning, and process evidence, students have incentives to develop the competencies that educators intend.

This article develops a design package for assessment integrity in AI-infused learning environments. The package contains two complementary outputs. First, we propose an evidence-centered assessment design framework that guides instructors from outcomes to evidence, policy boundary conditions, and verification moments. Second, we provide a rubric toolkit that operationalizes integrity through criteria such as reasoning quality, traceability, and AI-use disclosure, building on evidence about rubric design and formative use (Dawson, 2017; Panadero & Jonsson, 2013). Rather than prescribing a single policy stance, the package supports three common regimes: AI prohibited, AI permitted with constraints, and AI encouraged with transparency requirements. Across regimes, the principle is constant: assessment should elicit outcome-relevant evidence that the assessor can interpret with confidence.

2. Research Method

This study used an integrative synthesis and design science approach to produce practical assessment artifacts grounded in established theory and empirical evidence (Ruggiano & Perry, 2017). Integrative synthesis is appropriate when a problem spans multiple literatures and when the goal is to generate actionable design principles rather than estimate a single effect size (Cheong et al., 2023). We treated assessment integrity in AI-infused settings as a design problem: educators must specify what evidence of learning is required, anticipate plausible misuse pathways, and build assessment conditions that privilege ownership of reasoning (Felipe et al., 2025; Gonsalves, 2025).

Data sources included peer-reviewed literature on authentic assessment, constructive alignment, feedback and assessment for learning, academic integrity and contract cheating, and scholarship on generative AI in education and responsible tool use. We also reviewed guidance documents from higher education quality bodies and international organizations that address transparency, data protection, and responsible adoption. The screening process prioritized conceptual clarity and transferability across disciplines, focusing on sources that propose design principles, evaluation criteria, or implementation frameworks.

Analysis proceeded in three steps. Step 1 involved thematic coding of sources into design claims about what reduces misconduct opportunity and what strengthens the validity of assessment evidence (Kiger & Varpio, 2020; Lochmiller, 2021). Step 2 translated coded claims into a workflow of instructor decisions, expressed as an evidence-centered assessment design model with explicit inputs and outputs (Ilieva et al., 2025). Step 3 operationalized model decisions into rubric dimensions and example performance indicators, resulting in a matrix that can be adapted to local policy regimes (Brufau Alvira et al., 2025). To check usability, we conducted an expert review with experienced instructors and academic integrity practitioners in multiple disciplines, using structured feedback to refine wording and reduce ambiguity. Because the output is a design toolkit, the study emphasizes transparency of assumptions and boundaries rather than statistical generalization.

3. Result and Discussion

This section presents the Elevate Assessment Integrity Design framework and a rubric toolkit for authentic assessment in AI-infused learning environments. We first explain the evidence-centered workflow that connects learning outcomes, AI-use boundary conditions, and verification moments. We then provide rubric dimensions and transparency rules that can be adapted across disciplines and class sizes, supported by a model (Figure 1) and a matrix (Table 1).

3.1 Evidence-centered assessment design for AI-infused learning

Figure 1 presents the Elevate Assessment Integrity Design framework, which adapts evidence-centered design ideas to higher education assessment in the presence of generative AI. The framework begins with learning outcomes and constructive alignment: instructors specify the knowledge, skills, and dispositions that the task must evidence. From outcomes, the instructor derives an evidence claim, articulated as what the assessor must be able to see in order to judge mastery, such as argument quality grounded in sources, methodological justification, or debugging reasoning. This evidence claim names the observable features that signal learning, after which the framework addresses task format, AI-use boundary conditions, and verification mechanisms.

The focus on evidence claims rather than just topic descriptions represents a shift toward more resilient assessment structures. This design choice is consistent with recent frameworks that emphasize the need for assignments to promote academic integrity while remaining aligned with specific learning objectives (Felipe et al., 2025). Theoretically, this approach is grounded in the principles of constructive alignment and experiential learning, where the primary goal of assessment is to drive student learning behavior through authentic tasks (Salinas-Navarro et al., 2024).

The second decision in the framework is AI-use policy as boundary conditions rather than a binary stance. Boundary conditions clarify what AI is permitted to do—such

as brainstorming or language polishing—and what students must do themselves, such as producing the final structure and justifying choices. This move away from simple prohibition toward nuanced integration reflects a growing consensus that assessments should clarify how AI may be used as a tool within disciplinary practices (Gonsalves, 2025). Such boundary conditions are most effective when they prioritize the ethical adoption of AI and provide a clear scale for acceptable tool usage (Furze et al., 2024; Ilieva et al., 2025).

3.2 Rubric toolkit and transparency rules for acceptable AI use

The rubric toolkit operationalizes integrity by making the most valued criteria difficult to outsource and easy to evidence. Table 1 summarizes rubric dimensions that apply across common assessment types, including outcome-relevant reasoning, traceability, and disclosure of AI use. These dimensions are designed to ensure that work demonstrates disciplinary thinking rather than only fluent presentation.

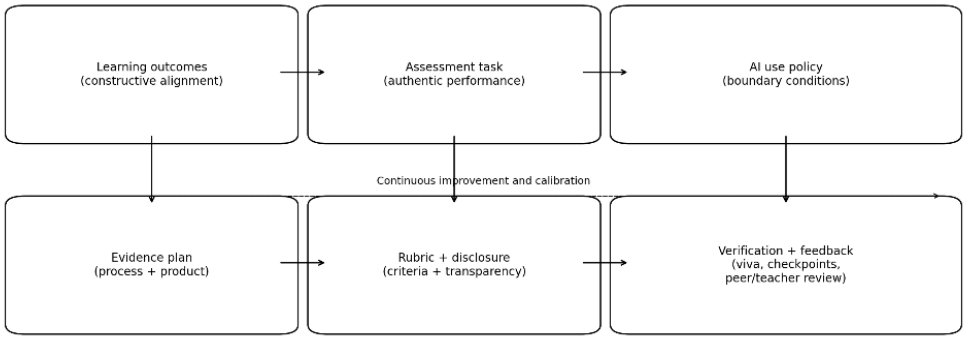

Figure 1. Elevate Assessment Integrity Design framework for authentic assessment in AI-infused learning.

By prioritizing process and reasoning over final outputs, this toolkit addresses the vulnerability of traditional written assessments to AI manipulation. Research indicates that authentic assessments alone cannot safeguard integrity if they remain focused on output; instead, they must rely on performative and process-based evidence (Kofinas et al., 2025). Disclosure of AI use is framed here as a part of scholarly transparency rather than a punitive measure. This approach aligns with recommendations to integrate AI into assessments to promote technical fluency and ethical reflection (Martin et al., 2025). Furthermore, requiring students to disclose their AI interactions encourages critical AI literacy and accountability (Perkins et al., 2023).

Table 1. Rubric toolkit dimensions and example indicators for transparent, integrity-supporting assessment.

Rubric dimension	Guiding question	Example indicators (adapt to discipline)
Outcome-relevant reasoning	Does the work demonstrate disciplinary thinking beyond surface fluency?	Claims are justified with evidence; reasoning steps are explicit; limitations are acknowledged.
Traceability and sourcing	Can key claims, data, or design choices be traced to sources or documented steps?	Citations match claims; calculations and data transformations are

		reproducible; versions or logs show iteration.
Process evidence and checkpoints	Is there credible evidence of development over time?	Drafts, notebooks, or checkpoint submissions demonstrate progression; feedback is incorporated with rationale.
AI-use disclosure	Is AI use documented clearly and honestly according to task rules?	Disclosure statement specifies tool roles (brainstorming, editing, debugging); student describes verification steps.
Verification readiness	Could the student explain and defend the work if asked?	Student can articulate choices, trade-offs, and errors; brief viva or in-class microtask aligns with submission.
Original contribution and personalization	Is the work meaningfully contextualized and not a generic template?	Uses local data, personal design decisions, or unique case constraints; examples are specific and relevant.
Ethical and responsible practice	Does the work respect academic norms, safety, and data governance?	Sensitive data are handled appropriately; claims avoid fabrication; model outputs are checked against sources.

Source: Processed by the researcher, 2026

A key design move in the toolkit is to separate language quality from reasoning quality. This ensures that students who use AI for writing polish do not receive an unfair advantage in demonstrating their mastery of core learning outcomes. This distinction is supported by studies on rubric validation, which suggest that criteria should specifically target the cognitive complexity of tasks to accurately measure student competence (Brufau Alvira et al., 2025). By allocating weight to justification and evidence selection, rubrics can support formative learning and critical engagement, especially when students are involved in the co-design of assessment expectations (Martin et al., 2025).

Implementation of this integrity-by-design approach also requires consideration of scalability and workload. While individualized verification like oral defense is ideal for small cohorts, large-scale courses may benefit from automated checkpoints or portfolio-based assessments that foreground iterative development (Kickbusch et al., 2025). Ultimately, the combination of Figure 1 and Table 1 provides a mechanism to move toward "integrity-by-design," where assessment systems are resilient to external outsourcing while fostering deeper student learning (Almpanis et al., 2025).

4. Conclusion

This article proposed a rights-respecting governance framework for learning analytics that translates ethical and legal principles into actionable controls, documentation artifacts, and an adoption roadmap. By organizing governance across the analytics lifecycle and emphasizing transparency and contestability, the framework supports institutions and vendors in implementing analytics as a support system rather than a surveillance apparatus.

The framework is intentionally pragmatic. It provides a principle-to-control mapping and a maturity model that can be used for institutional policy, procurement due diligence, and program evaluation. Future work should empirically evaluate the framework in diverse institutional contexts, including resource-constrained universities and cross-border EdTech arrangements, and should develop measurement tools for learner trust and perceived legitimacy of analytics interventions. Ultimately, learning

analytics will be sustainable only if it remains legitimate in the eyes of learners and the public. Rights-respecting governance provides a path to that legitimacy by embedding accountability, intelligibility, and due process into the everyday routines of data-intensive education.

References

Almpanis, T., Conroy, D., & Joseph-Richard, P. (2025). Practical implications of generative AI on assessment: Snapshot of early reactions to assessment redesign in an HRM and a psychology course. Journal of Learning Development in Higher Education. https://doi.org/10.47408/jldhe.vi34.1307

Ardito, C. G. (2023). Contra generative AI detection in higher education assessments. ArXiv. https://doi.org/10.48550/arxiv.2312.05241

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. Journal of Learning Development in Higher Education. https://doi.org/10.47408/jldhe.vi34.1307

Cheong, H.-I., Lyons, A., Houghton, R., & Majumdar, A. (2023). Secondary qualitative research methodology using online data within the context of social sciences. International Journal of Qualitative Methods. https://doi.org/10.1177/16094069231180160

Felipe, A. L., Khwakhali, U. S., & Nguyen, T. N. (2025). A framework for assessment design in the era of generative AI: Case study of take-home assignment in software-related courses. 2025 10th International STEM Education Conference (iSTEM-Ed). https://doi.org/10.1109/istem-ed65612.2025.11129352

Francis, N. J., Jones, S., & Smith, D. P. (2025). Generative AI in higher education: Balancing innovation and integrity. Journal of Learning Development in Higher Education. https://doi.org/10.47408/jldhe.vi34.1307

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI supported assessment. Journal of Learning Development in Higher Education. https://doi.org/10.47408/jldhe.vi34.1307

Gonsalves, C. (2025). Contextual assessment design in the age of generative AI. Journal of Learning Development in Higher Education. https://doi.org/10.47408/jldhe.vi34.1307

Ilieva, G., Yankova, T., Ruseva, M., & Kabaivanov, S. (2025). A framework for generative AI-driven assessment in higher education. Information, 16(6), 472. https://doi.org/10.3390/info16060472

Kickbusch, S., Ashford-Rowe, K., Kemp, A., Boreland, J., & Huijser, H. (2025). Beyond detection: Redesigning authentic assessment in an AI-mediated world. Journal of Learning Development in Higher Education. https://doi.org/10.47408/jldhe.vi34.1307

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. Medical Teacher, 42(8), 846-854. https://doi.org/10.1080/0142159X.2020.1755030

Kofinas, A. K., Tsay, C., & 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(1). https://doi.org/10.1111/bjet.13585

Lochmiller, C. R. (2021). Conducting thematic analysis with qualitative data. The Qualitative Report, 26(6), 2029-2044. https://doi.org/10.46743/2160-3715/2021.5008

Martin, A. F., Tubaltseva, S., Harrison, A., & Rubin, G. J. (2025). Participatory co-design and evaluation of a novel approach to generative AI-integrated coursework assessment in higher education. Behavioral Sciences, 15(6), 808. https://doi.org/10.3390/bs15060808

Perkins, M., Roe, J., Postma, D., McGaughran, J., & Hickerson, D. (2023). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. Journal of Learning Development in Higher Education.https://doi.org/10.47408/jldhe.vi34.1307

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. Journal of Learning Development in Higher Education. https://doi.org/10.47408/jldhe.vi34.1307

Perkins, M., Roe, J., Postma, D., McGaughran, J., & Hickerson, D. (2023). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. Journal of Learning Development in Higher Education. https://doi.org/10.47408/jldhe.vi34.1307

Ruggiano, N., & Perry, T. E. (2017). Conducting secondary analysis of qualitative data: Should we, can we, and how? Qualitative Social Work, 18(1), 81-97. https://doi.org/10.1177/1473325017700701.

Salinas-Navarro, D., Vilalta-Perdomo, E., Michel-Villarreal, R., & Montesinos, L. (2024). Using generative artificial intelligence tools to explain and enhance experiential learning for authentic assessment. International Journal of Educational Technology in Higher Education. https://doi.org/10.1186/s41239-024-00462-2