International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(13s) 2026 pp. 860-867
https://doi.org/10.70917/ijcisim-2026-4130
Article

From Policy to Practice: Academic Leaders' and Lecturers' Experiences of Implementing Generative AI Governance and AI-Enabled Assessment Reform in Higher Education

Zhou Dewen[1], Fajar Rezeki Ananda Lubis[2], Christin Agustina Purba[3]

[1,2,3]Universitas Prima Indonesia
Corresponding Author: Fajar Rezeki Ananda Lubis,
Email: fajarrezekiananda@gmail.com

Abstract: Generative artificial intelligence has compelled universities to render generalized promises of integrity, innovation, privacy, and fairness into working guidelines of teaching and evaluation. This secondary qualitative synthesis explores the experiences of academic leaders and lecturers with this translation, what assessment reforms are surfacing, and what organisational conditions should work. Peer-reviewed higher education research published between 2016 and July 2026 was systematically searched. To maintain analytical independence, sources were not used to create the Literature Review. The results were based on a limited number of 15 primary empirical studies published between 2024 and 2026, all in the last five years. Three themes were generated using reflexive thematic synthesis. First, governance is established based on negotiated local discretion; however, discretion is inequitable if common procedural protections are fragile. Second, the shift of assessment reform is chiefly in the direction of examining processes, which can be seen, rather than assessing finished work, such as staged work, oral explanation, reflective disclosure, contextual tasks, and critical assessment of AI output. Third, sustainable implementation requires tuned trust, discipline-related professional learning, approved infrastructure, moderation, and appreciation of redesign workload. The synthesis questions both the prohibition-led governance and the unquestioning adoption. This suggests a tiered architecture where minimum safeguards are specified by institutions, programs understand the requirements of the discipline, and modules declare task-level permissions and evidence. The evidence is still limited in terms of sample size, self-reporting, single institutions, and limited longitudinal assessments. Therefore, the validity of learning, equity, workload, transparency, and procedural justice should be used to evaluate governance, not policy publication or technology acceptance alone.

Keywords: generative artificial intelligence; higher education governance; academic leadership; lecturers; AI-enabled assessment; academic integrity; policy enactment

1. Introduction

The free release of ChatGPT in late 2022 transformed a technology problem under development into a direct governance threat to postsecondary education. Generative artificial intelligence (GenAI) can write, render translation, summarise, code, simulate conversation, and feedback generation scaleable enough to disrupt both the traditional understanding of authorship and student performance. The institutional challenge does not pertain to the use of a tool by students or staff. It involves the way in which universities translate pledges to academic integrity, educational innovation, privacy, accessibility, and academic freedom into decisions that can be comprehended, implemented, and justified in disciplines and examinations. Policy analysis indicates that institutional standpoints, acceptable uses, reporting regulations, and resources accessible to staff vary significantly [1-5].

Initial reactions tended to focus on restriction, detection, or caution. This kind of response could be justified since an unmonitored written product could no longer be handled as a matter of course as crystal-clear evidence of personal learning. However, an enforcement-based government is fragile. Detection is most likely probabilistic,

functional embedded AI is increasingly opportune to ban, and unsubstantiated suspicion may create procedural injustice. An equally lenient reaction is also unhealthy, as GenAI can replace intellectual labour, reveal trade secrets, create prejudice, and contribute to imbalances in access and digital literacy. Faculty-informed governance then constitutes engagement, disciplinary interpretation, and clear safeguards and not a set universal rule [6].

The key point in this problem is assessment reform. Assessment validity is impaired when a plausible final product can be created without undertaking the desired learning. Some suggested reactions are staged submissions, oral descriptions, live performances, reflective disclosures, critiques of the generated work, and supervised components [7-13]. Such measures might enhance the evidence of reasoning, but they might also lead to more workload and accessibility barriers and diminish comparability among students'scores. This study sought to answer the following questions: How is the institutional GenAI policy understood and implemented by academic leaders and lecturers? What are the emerging assessment reforms? In what organisational conditions can sustainable implementation be realised or restricted? A secondary qualitative synthesis divides background scholarship and the current primary Findings corpus before comparing the two sets of evidence in the Discussion section.

2. Literature Review

2.1 Governance as Policy Enactment

GenAI governance is commonly formulated as a drafting problem, as though language can be formulated to achieve consistent practice. Comparative policy research suggests that institutional texts outline broad lines, but local actors determine those lines in relation to specific undertakings. The policy education framework suggested by Chan [1] suggests that there is a connection between institutional, curriculum, and assessment decisions; however, its normative scope does not define the ways in which conflicting values are sorted out in times of tension. Similarly, Luo [2] demonstrated that even if GenAI ensures that authorship becomes distributed, assessment policies commonly remain relatively close to a rather limited interest in originality. This is a significant criticism, and the liberalisation of originality without designing anyone to make intellectual contributions culpable might merely render standards less progressive than more open.

Another gap in implementation was determined by cross-institutional analyses. Wang et al. [3] found that university policies and resources varied in the scope and details of their operations, whereas Sutedjo et al. [4] found that the faculty-facing provision had a higher degree of information than long-term redesign support. Ganguly et al. [5] revealed alignment in transparency, authorship, privacy, and external obligations in research guidance. The studies show that there is institutional signalling; however, documentary and web-based design cannot answer the question of whether rules are perceived, resources, or practiced. Dotan et al. [6] consider legitimacy more directly, stating that legitimacy should be influenced by faculty deliberation based on the concept that academic freedom, disciplinary difference, educational purpose, and participation should influence adoption. This strategy enhances procedural legitimacy but lacks self-selected participation and has no outcome measures undermining effectiveness claims.

Collectively, the literature considers governance as a negotiated institutional practice rather than mechanical compliance. The distribution of accountability remains unsolved. Proportionality and disciplinary relevance may be enhanced through local discretion; however, this may hand over to the individual lecturers the legal, relational, and workload risks to the institution. Thus, there should be a credible system which differentiates between authorised adaptation and unsupported improvisation.

2.2 Assessment Validity, Integrity, and Redesign

The literature on assessment finds less and less motivation to make tasks resistant to AI and more motivation to render learning and judgement visible. Bower et al. [7] identified the support of educators for authentic assessment, AI literacy, ethical teaching, and critical analysis of generated material. Similar priorities were translated into redesign principles by Lye and Lim [8], and Weng et al. [9] and Zhao et al. [10] demonstrated that the evidence base is dominated by conceptual discussion, perceptions, and small interventions as opposed to solid outcome evaluation. This makes the reform agenda plausible pedagogically, but empirically immature.

Academic integrity scholarship also reveals the shortcomings of a detection-based model. Bittle and El-Gayar [11] discovered long-standing anxiety with respect to misconduct, authorship, and inconsistency in policies in the research domain. According to Cotton et al., the design of assessment and student education should be used to complement the control of integrity [12]. The Artificial Intelligence Assessment Scale created by Perkins et al. [13] provides task-based permission on a graduated basis, which is clearer than its opposite, blanket prohibition. However, a scale can be applied procedurally without the need to test the potential assessment of the resulting measurement to

determine whether the intended capability is measured. General educational critique cautions that advantages such as response and customisation are together with hallucinating, favouritism, addictiveness, and disproportionate access [14-15]. Recent research on perception reveals that there are no consistent limits between what employees and students regard as reasonable help and what is considered malpractice [16-17].

Therefore, there is no inherently valid or just assessment form. Oral defence can be advantageous to those students who can argue, but it puts others at a disadvantage. Customised assignments can also minimise outsourcing; however, the moderation process becomes more challenging. Repeat checkpoints can reinforce process evidence and add to surveillance and marking. Reform needs to be balanced with positive alignment, ease of use, reliability, workload, due process, and not just novelty or seeming opposition to AI.

2.3 Theoretical Framework

This synthesis is a mix of policy enactment, constructive alignment and implementation capacity. Institutional policy-making views institutional regulations as construed and translated by actors in a situation. Guidance is read by academic leaders and lecturers based on disciplinary standards, the history of the programme, students' needs, professional control, and resources at hand. This view clarifies why the same policy wording can produce variant rules of modules and why ambiguity can provide a chance of innovation in one environment but a state of inequity in another [1-6].

Constructive alignment measures the consistency between the planned learning outcomes, instructional activity, and allowed AI utilisation, as well as the coherence between assessment evidence. GenAI does not nullify an assessment simply because there is one. A threat to validity occurs when a student can generate the assessed output but cannot show the rationale or practice intended [7-13]. Implementation capacity introduces the organisational requirements to maintain aligned practice, such as sanctioned infrastructure, employee wisdom, workload distribution, moderation, privacy, accessibility skills, and feedback. The two-part framework is stronger than the frontiers of technology acceptance because goodwill to use GenAI does not provide educational legitimacy, evidential validity, or institutional feasibility.

2.4 Research Gaps

Five gaps remain: first, the broader discipline which favours student attitudes, and deans, department heads, quality heads, assessment teams, and precariously employed lecturers are all underrepresented. Second, there is much literature that documents intentions or initial perceptions and not actual performance over time. Third, policy document studies seldom proceed into classroom judgement, misbehaviour policies, appeals, or redesigns [3-5]. Fourth, learning, reliability, accessibility, equity, and staff workload are rarely measured collectively in assessment studies [9-13]. Fifth, institutions of the Global North and research are overrepresented, undermining transfer to vocational and resource-limited Global South environments.

Government and assessment research are also divided by a conceptual gap. Principles and risks are studied in governance and local task design in pedagogical studies. This separation hides how procurement, workload, professional development, moderation, and due process determine assessment validity. The current synthesis bridges this gap by conceptualising governance and assessment reform as a single implementation issue and comparing the established body of literature to a specific set of recent primary works.

3. Methods

A methodical secondary qualitative synthesis was performed. No new participant data were obtained. The search was structured based on a combination of generative AI, ChatGPT, higher education, university, governance, policy, academic integrity, leadership, lecturer, faculty, assessment, redesign, workload, trust, and professional development. This study considered peer-reviewed publications published between 2016 (January) and 2026 (July). The Literature Review relied on both conceptual and review, as well as policy and mainstream empirical scholarship on establishing the field. The results were limited to first-order empirical studies conducted within the last five years, i.e., 20212026. Qualitatively, the 15 eligible findings publications were published in 2024-2026 since evidence of credible post-ChatGPT implementation practices was concentrated in that period.

The findings had eligibility criteria of identifiable empirical design, higher educational environment, and findings of academic leaders, lecturers, teaching team, or those who participated in the conduct of governance or the implementation of assessment. Eligible studies included interviews, focus groups, surveys with qualitative aspects, phenomenography, institutional case studies, participatory co-design, and practitioner autoethnography. The findings

did not include reviews, conceptual papers lacking data, preprints, student-only studies, or document audits. To avoid the possibility of recycling background evidence as findings, all the studies mentioned in the Literature Review were not included in the findings corpus.

The review was based on an interpretation with a thematic synthesis guided by advice on reflexive thematic analysis and qualitative evidence synthesis [18-20]. The study's characteristics and limitations were documented, and the results were coded for interpretation as policy, discretion, assessment response, trust, capability, workload, equity, and organisational support. Codes were matched in national settings and methods, grouped together in candidate themes, and retained only where multiple independent studies supported the topic. The claims that it entails are not about pooled effects or cause and effect but about recurrent meanings and patterns of implementation. The transparency of the methods used to judge quality included the adequacy of sampling, close association with the implemented practice, context detail, triangulation, and explicit constraints. The use of published evidence instead of engaging in conducting an ethical approval was not necessary.

4. Findings

4.1 Theme 1: Governance Is Distributed, Negotiated, and Uneven

More recent primary data demonstrate that GenAI policy finds its way into practice via local interpretation and not direct compliance. In a study at an international university in the United Arab Emirates, Alsharefeen and Al Sayari [21] discovered that faculty members doubted the transparency and specificity of integrity policy and tended to discuss, review drafts, warn, and revise instead of reporting. This latitude can cushion relationships and permit proportional responses; however, it might create dissimilar consequences in the event of similar cases relying on personal confidence or endurance.

The institutional level of leadership studies portrays the same trend. Romanian university heads characterised the existence of structures based on formal AI groups and hubs as less organised guidance, alongside placing themselves in the spectrum of regulation, innovation, and academic autonomy [22]. An example of governance through consultation, comparison, and iterative learning of an institution came from a South African case of task teams, where they did not engage in the adoption of an external template but instead learnt through iterative learning [23]. The two studies advocate context-sensitive policy, though leadership and insider accounts can exaggerate coherence and under-represent lecturers who find implementation discordant.

The difference between distributed governance and policy abandonment is supported by broader evidence from educators. Lee et al. [24] discovered that educators considered institutional direction and professional support required responsible teaching, whereas Palestinian faculty reported opportunities and policy, infrastructure and ethical limitations [25]. Luo et al. [26] also demonstrated that university educators perceived assessment as a dilemma space where the maximisation of integrity, learning, feasibility and fairness were not all possible at the same time. In this context, local discretion seems inevitable. It relies on generic procedural floors of accepted tools, privacy, disclosure, evidence, advice, reporting, and appeal. Institutions seem to be flexible without such floors, and risk is externalised to individual staff.

4.2 Theme 2: Assessment Reform Makes Process and Judgement Visible

The second theme regards the movement of the final products of policing towards evidencing the production of learning. Farazouli et al. [27] established that university professors reevaluated home examinations once they had tested the quality of things that AI chatbots were able to produce, and they themselves doubted whether standard assignments continued to serve defensible judgement. Tran et al. [28] also reported interest in continuous assessment, contextualised problems, active learning, and non-written parts. These proposals were associated with course design, as opposed to detection; however, much of the evidence was about building practice, as opposed to quantifiable long-term impacts.

There is more direct implementation research which demonstrates the functioning of process visibility. In the participatory co-design applied by Martin et al. [29] to the postgraduate psychology programme, explicit AI-use conditions, critical reflection, blinded marking, and criteria were developed to address both disciplinary performance and AI literacy. Dosumu et al. [30] reported the reaction of an accounting marking team to uncertainty on an open-book online test, which eventually switched the focus to determining the AI impact on whether the task was valid and required common marking guidelines. Both articles indicate that redesigning is not an easy rearrangement of instructions but an organisational activity. Scalability is also uncertain because of their small and localised samples.

Faculty studies in South Africa found possible applications of formative feedback, generation of questions, generation of marking assistance, and generation of critiques of model output, as well as a possible overreliance and lack of clear policy [31]. Van den Berg and Papadopoulos [32] observed varied acceptance patterns in teachers and students in AI-based summative assessment, which is a lesson for institutions that technical availability does not generate common legitimacy. Luo et al. [26] demonstrated the reason teachers tend to mix the answers but not to pick one of the answers. All the instances, taken together, show the staged drafts, oral explanation, reflective disclosure, authentic performance, contextual judgement, and critical analysis of AI output. Nevertheless, process evidence might turn performative in itself, oral elements might pose obstacles to accessibility, and high personalisation might undermine comparability. Reasonable adjustments and moderation are necessary.

4.3 Theme 3: Capacity, Trust, and Workload Determine Sustainability

The third theme is that execution requires less zeal than group abilities. Ellis et al. [33] found qualitatively distinct teacher experiences on a continuum of learning-based engagement to suspiciousness. Positive use was conditional upon verification, fitting work into tasks, and conceptions of teaching that tended to remove the binary perspective of staff as adopters or resisters. Lyu et al. [34], also managed to discover that trust and distrust were different dimensions among instructors of the United States. Calibrated reliance, even without using a model, should thus be the focus of professional learning, instead of merely raising the level of acceptance.

In Linden et al. [35], teaching the staff was a matter of trial and error. They had the possibility of imagining change in pedagogy but had gaps in capability and moral confusion, as well as practical limitations. Such strain between perceived educational value and perceived anxiety about accuracy, integrity, and preparedness has also been reported by Lee et al. [24]. Omar et al. [25] found these tensions in the conditions of resources and infrastructure, which demonstrated that the attitudes could not be discussed in isolation of the institutional capacity. The implication of these studies is that isolated and infrequent workshop sessions and sound lessons alone are not enough, since responsible use entails recurring disciplinary judgement.

Workload links individual experiences with governance. According to Tran et al. [28], formative, contextual, and dialogic designs involve more reviews and interactions. Dosumu et al. [30] demonstrated that uncertain cases augment the coordination requests of markers, whereas Martin et al. [29] necessitated co-design, rubric development, and assessment. Alsharefeen and Sayari [21] correlated policy weakness with inconsistent integrity practices. Throughout the corpus, a lack of time to invest in professional development was a danger of being solely symbolic. To support sustainable reform, approved tools, discipline-based exemplars, assessment design support, moderation arrangements, accessibility advice, and redesign labour recognition are required. Capacity is also distributive: individual experimentation benefits staff with time and technical confidence, and those whose staff are contingent and in large-enrolment modules have a smaller number of viable options.

5. Discussion

The results support the statement made in the Literature Review that policy is not enacted but implemented and reveal a more pronounced difference between legitimate participation and uncontrolled inconsistency. Chan [1] and Dotan et al. [6] focused on multi-level governance and faculty-informed governance. Theme 1 demonstrates how such principles work under the influence of life itself. The contextual interpretation is facilitated by faculty discretion in the United Arab Emirates [21], task-team work with iteration in South Africa [23], and leadership boundary work in Romania [22]. However, they also demonstrated that consultation cannot be used as a certain procedure. The current synthesis thus adds to the literature on policy scholarship by proposing that the freedom of academic institutions requires an institutional procedural floor. Not to be advisory, documented, and appealable is not to be empowered; it is risk transfer.

These findings further enhance the documentary evidence. Policies, resources, and frequent types of risks were identified by Wang et al. [3], Sutedjo et al. [4], and Ganguly et al. [5]. Theme 3 justifies why information provision is not valuable in terms of implementation: Differences in the experience of teachers [33], the calibration of trust [34], and trial and error experimentation [35] prove that the staff members require disciplinary exemplification, facilitated practice, and being allowed to apply critical rejection. The assessment redesign and marking coordination workload evidence [28-30] also helps reposition time allocation as a governance process. Any policy in which process-rich assessment is demanded but there is no change to workload models is internally contradictory.

Theme 2, in wide terms, is in line with the focus of the Literature Review on authenticity, AI literacy, and graduated permissions [7-8,13]. The Findings supply mechanisms and tensions. Permitted use can be assessed through

the use of co-designed conditions and reflective criteria [29], and teachers' responses to home examination and summative assessment reveal that legitimacy needs to be negotiated, not assumed [27,32]. Optimistic redesign claims also qualify for this comparison. Promising directions have been reported in reviews, but with little evidence of outcomes [9-11]. The main studies are small, self-reported, and preliminary; thus, process-visible evaluation would only be a hypothesis to be tested and not a cure-fit solution.

The synthesised theoretical framework elucidates one of the main contradictions in the literature. Local flexibility is needed because of policy enactment, constructive alignment because differentiation at the task level is pedagogical in nature, and implementation capacity because excessive localisation constitutes an organisational hazard. The layered architecture provides a justifiable solution. Institutions must establish privacy, procurement, disclosure, accessibility, due process, and minimum integrity requirements. These requirements are to be interpreted by programs against disciplinary and professional standards. Modules must declare the banning, optional, or compulsory character of AI, which should list the places of acceptable application and what the evidence of reasoning should be. All three layers should be corrected using moderation data, cases, appeals, student results, and staff workload. This kind of feedback also eliminates the possibility of either extreme of governance becoming either hardcore centralisation or creasy-casy individual discretion.

The synthesis is also tentative in this study. The 15 original works cover a variety of nations and methods, but are mostly based on small samples, self-reports, single institutions, or early versions of their models. Few studies have tracked policy over time, made comparisons between learning in pre- and post-redesign groups, or experimented on the differences in impact on disabled, multilingual, economically disadvantaged, or digitally excluded students. This indicates progress rather than a predetermined pattern.

6. Conclusion

The exercise of GenAI governance as part of education can only be educationally relevant if it is executed by assessment, professional judgment, and the support of an institution. Academic executives and instructors bargain policy boundaries, redefine integrity, redesign jobs, and internalise the risks brought about by uncertainty. The following three conclusions can be drawn from this study. Government is dispersed and arbitrary; believable change is transparent in its influence on learning processes and judgement; and sustainability lies in judicious trust, pooled capacity, and recognised work. Detection and prohibition will never be able to provide a secure base for reliable assessment, and in the case of blind adoption, there is a misunderstanding of the availability and educational merit. An approach to be defended additionally features ordinary procedural protections, discipline-sensitive design, participatory revision, learning, fairness, workload, and due process assessment.

In the case of institutional leaders, the focus is on eliminating a single-time policy publication and moving to a reviewable system of governance. Universities ought to establish minimum standards regarding tools that may be used, privacy, disclosure, evidence, reporting, reasonable adjustment, and appeal, and should permit programs to provide interpretations of disclosures regarding disciplinary matters. Lecturers, students, professional services, librarians, disability specialists, and contingent staff should be included in the governance groups. Data on cases and appeals must be tracked in cases where the outcomes are inconsistent or discriminatory.

Program teams and lecturers redesign must start with desired learning as opposed to the purpose of conquering a tool. Each task must indicate the prohibition, optionality, or necessity of GenAI; what uses are permitted; what is treated, and how it is disclosed; and what evidence is conveyed in support of judgement. Only when compatible, available, and commensurate with one another, staged work, oral explanation, authentic performance, and critical AI evaluation should be chosen. Dilution must be applied to defend comparability by markers and student groups.

Generic awareness sessions are insufficient for staff development. Institutions should finance discipline-based redesign clinics, communities of practice, exemplars, technical support, and time out of workload models. It should be developed to deal with verification, hallucination, bias, privacy, access, record and output of prompts, and prompt limitations of detection. It is oriented towards calibrated professional judgement but not unthinking adoption.

In research, priorities are longitudinal and comparative, adhering to policy during design and enforcement to classroom implementation and revision. Assessment should incorporate leaders, student-faculty assessment boards, contingent faculty and professional staff, underrepresented settings, student learning, reliability, accessibility, staff time, integrity cases, and appeals. This would require evidence to conclude that GenAI governance yields more trustworthy and fairer higher education as opposed to increased acceptance.

Acknowledgments

None.

Funding

This research received no external funding.

Conflict of Interest

The author declares no conflict of interest.

Data Availability Statement

No new data were created or analysed. This study synthesised published research.

References

1. Chan CKY. A comprehensive AI policy education framework for university teaching and learning. Int J Educ Technol High Educ. 2023;20:38. https://doi.org/10.1186/s41239-023-00408-3.
2. Luo J. A critical review of GenAI policies in higher education assessment: a call to reconsider the originality of students' work. Assess Eval High Educ. 2024;49(5):651-664. https://doi.org/10.1080/02602938.2024.2309963.
3. Wang H, Dang A, Wu Z, Mac S. Generative AI in higher education: seeing ChatGPT through universities' policies, resources, and guidelines. Comput Educ Artif Intell. 2024;7:100326. https://doi.org/10.1016/j.caeai.2024.100326.
4. Sutedjo A, Liu SP, Chowdhury M. Generative AI in higher education: a cross-institutional study on faculty preparation and resources. Stud Technol Enhanc Learn. 2025;4(1). https://doi.org/10.21428/8c225f6e.955a547e.
5. Ganguly A, Johri A, Ali A, McDonald N. Generative artificial intelligence for academic research: evidence from guidance issued for researchers by higher education institutions in the United States. AI Ethics. 2025;5(4):3917-3933. https://doi.org/10.1007/s43681-025-00688-7.
6. Dotan R, Parker LS, Radzilowicz J. Responsible adoption of generative AI in higher education: developing a points to consider approach based on faculty perspectives. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency. New York: Association for Computing Machinery; 2024. p. 2033-2046. https://doi.org/10.1145/3630106.3659023.
7. Bower M, Torrington J, Lai JWM, Petocz P, Alfano M. How should we change teaching and assessment in response to increasingly powerful generative artificial intelligence? Outcomes of the ChatGPT teacher survey. Educ Inf Technol. 2024;29:15403-15439. https://doi.org/10.1007/s10639-023-12405-0.
8. Lye CY, Lim L. Generative artificial intelligence in tertiary education: assessment redesign principles and considerations. Educ Sci. 2024;14(6):569. https://doi.org/10.3390/educsci14060569.
9. Weng X, Xia Q, Gu M, Rajaram K, Chiu TKF. Assessment and learning outcomes for generative AI in higher education: a scoping review on current research status and trends. Australas J Educ Technol. 2024;40(6):37-55. https://doi.org/10.14742/ajet.9540.
10. Zhao J, Chapman E, Sabet PGP. Generative AI and educational assessments: a systematic review. Educ Res Perspect. 2024;51:124-155. https://doi.org/10.70953/ERPv51.2412006.
11. Bittle K, El-Gayar O. Generative AI and academic integrity in higher education: a systematic review and research agenda. Information. 2025;16(4):296. https://doi.org/10.3390/info16040296.
12. Cotton DRE, Cotton PA, Shipway JR. Chatting and cheating: ensuring academic integrity in the era of ChatGPT. Innov Educ Teach Int. 2024;61(2):228-239. https://doi.org/10.1080/14703297.2023.2190148.
13. Perkins M, Furze L, Roe J, MacVaugh J. The Artificial Intelligence Assessment Scale (AIAS): a framework for ethical integration of generative AI in educational assessment. J Univ Teach Learn Pract. 2024;21(6). https://doi.org/10.53761/q3azde36.
14. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ. 2023;103:102274. https://doi.org/10.1016/j.lindif.2023.102274.
15. Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R, et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learn Environ. 2023;10:15. https://doi.org/10.1186/s40561-023-00237-x.
16. Karkoulian S, Sayegh N, Sayegh N. ChatGPT unveiled: understanding perceptions of academic integrity in higher education, a qualitative approach. J Acad Ethics. 2025;23:1171-1188. https://doi.org/10.1007/s10805-024-09543-6.
17. Ardoin PJ, Hicks WD. Fear and loathing: ChatGPT in the political science classroom. PS Polit Sci Polit. 2024;57(4):583-593. https://doi.org/10.1017/S1049096524000131.
18. Braun V, Clarke V. One size fits all? What counts as quality practice in reflexive thematic analysis? Qual Res Psychol. 2021;18(3):328-352. https://doi.org/10.1080/14780887.2020.1769238.
19. Snyder H. Literature review as a research methodology: an overview and guidelines. J Bus Res. 2019;104:333-339. https://doi.org/10.1016/j.jbusres.2019.07.039.
20. France EF, Cunningham M, Ring N, Uny I, Duncan EAS, Jepson RG, et al. Improving reporting of meta-ethnography: the eMERGe reporting guidance. BMC Med Res Methodol. 2019;19:25. https://doi.org/10.1186/s12874-018-0600-0.

21. Alsharefeen R, Al Sayari N. Examining academic integrity policy and practice in the era of AI: a case study of faculty perspectives. Front Educ. 2025;10:1621743. https://doi.org/10.3389/feduc.2025.1621743.
22. Cernicova-Buca M, Palea A. Leading through transformation: university responses to generative AI in Romanian higher education. Prof Commun Transl Stud. 2026;19:44-55. https://doi.org/10.59168/ddjk2862.
23. Thaldar D, Botes M, Badru A, Chenia H, Duma S, Dlamini SB, et al. Generative AI governance in higher education: a case study from Africa. Front Polit Sci. 2025;7:1666661. https://doi.org/10.3389/fpos.2025.1666661.
24. Lee D, Arnold M, Srivastava A, Plastow K, Strelan P, Ploeckl F, et al. The impact of generative AI on higher education learning and teaching: a study of educators' perspectives. Comput Educ Artif Intell. 2024;6:100221. https://doi.org/10.1016/j.caeai.2024.100221.
25. Omar A, Shaqour AZ, Khlaif ZN. Attitudes of faculty members in Palestinian universities toward employing artificial intelligence applications in higher education: opportunities and challenges. Front Educ. 2024;9:1414606. https://doi.org/10.3389/feduc.2024.1414606.
26. Luo J, Keung CPC, Tang HHH. Assessment as a dilemmatic space in the GenAI age: mapping and unpacking university teachers' conflicting priorities in assessment. Assess Eval High Educ. 2025;50(4):607-621. https://doi.org/10.1080/02602938.2024.2444890.
27. Farazouli A, Cerratto-Pargman T, Bolander-Laksov K, McGrath C. Hello GPT! Goodbye home examination? An exploratory study of AI chatbots' impact on university teachers' assessment practices. Assess Eval High Educ. 2024;49(3):363-375. https://doi.org/10.1080/02602938.2023.2241676.
28. Tran TM, Bakajic M, Pullman M. Teacher's pet or rebel? Practitioners' perspectives on the impacts of ChatGPT on course design. High Educ. 2025;90(3):777-800. https://doi.org/10.1007/s10734-024-01350-7.
29. Martin AF, Tubaltseva S, Harrison A, Rubin GJ. Participatory co-design and evaluation of a novel approach to generative AI-integrated coursework assessment in higher education. Behav Sci (Basel). 2025;15(6):808. https://doi.org/10.3390/bs15060808.
30. Dosumu O, Porumb VA, Stafford A, Zimmer A. In the wake of ChatGPT: early reflections on marking open-book online accounting assessments. Account Educ. 2026;35(4):425-456. https://doi.org/10.1080/09639284.2025.2487487.
31. Maistry SM, Singh UG. Faculty perspectives on the role of ChatGPT-4.0 in higher education assessments. J Educ. 2025;(98):86-102. https://doi.org/10.17159/2520-9868/i98a05.
32. van den Berg S, Papadopoulos PM. Summative assessment with artificial intelligence: qualitative analysis and comparison of technology acceptance in student and teacher populations. Innov Educ Teach Int. 2025;62(5):1529-1544. https://doi.org/10.1080/14703297.2024.2436613.
33. Ellis R, Han F, Cook H. Qualitatively different teacher experiences of teaching with generative artificial intelligence. Int J Educ Technol High Educ. 2025;22:33. https://doi.org/10.1186/s41239-025-00532-2.
34. Lyu W, Zhang S, Chung T, Sun Y, Zhang Y. Understanding the practices, perceptions, and (dis)trust of generative AI among instructors: a mixed-methods study in U.S. higher education. Comput Educ Artif Intell. 2025;8:100383. https://doi.org/10.1016/j.caeai.2025.100383.
35. Linden T, Yuan K, Mendoza A. The potential and challenges of integrating generative AI in higher education as perceived by teaching staff: a phenomenological study. Inf Syst Educ J. 2025;23(3):16-30. https://doi.org/10.62273/IENP8578.