Id: 1885223

Participatory Educational Research (PER)
Vol.13(4), pp. 290-311, July 2026
Available online at http://www.perjournal.com
ISSN: 2148-6123
http://dx.doi.org/10.17275/per.26.60.13.4

Designing AI-Aware Assessment in Higher Education: A Cross-National Analysis of Generative AI Guidance and Design Patterns in English-Medium Universities

Emrah Baki Başoğlu*
School of Foreign Languages, Zonguldak Bulent Ecevit University, Zonguldak, Türkiye
ORCID: 0000-0002-5146-0440

Article history
Received:
09.02.2026

Received in revised form:
21.03.2026

Accepted:
24.04.2026

Key words:
higher education assessment; generative AI; document analysis; assessment design patterns; academic integrity policy

Generative AI is reshaping coursework and assessment in higher education, yet institutional guidance still varies in how clearly broad principles are translated into assessment practice. Guided by pattern-based educational design, the study used a framework-based document analysis of 21 official, public-facing institutional and teaching-unit guidance documents from universities in Europe, North America, Australasia, and Asia. The analysis combined directed qualitative coding with descriptive cross-case summaries of permitted and prohibited uses, disclosure and verification requirements, ethical safeguards, and ratings of clarity, actionability, and assessment specificity. Assessment-related configurations were synthesised into recurrent design patterns, and a descriptive ethical gap map compared the presence or absence of provisions related to privacy, consent, retention, fairness, and accessibility. Across the corpus, all universities allowed some student use of generative AI under conditions, and almost all required disclosure. Policies were generally clear, actionable, and assessment-specific. Four patterns were especially prominent: (1) process-evidence portfolios with AI-use logs, (2) AI-assisted drafting plus in-class or oral verification, (3) critical engagement with AI outputs, and (4) secure high-stakes exams combined with AI-enabled coursework. The ethical mapping showed stronger attention to public-upload risks and bias awareness than to data retention, consent language, and accessibility alternatives. Overall, the findings suggest that universities in this purposive sample have moved beyond blanket prohibition toward more assessment-aware guidance, while important governance and equity gaps remain. The resulting pattern library is offered as a provisional, policy-grounded resource for translating high-level AI principles into concrete assessment design options.

Introduction

Generative AI and higher education assessment

The rapid emergence of generative artificial intelligence (GenAI) in late 2022 has unsettled established assumptions about coursework and assessment in higher education. GenAI is now embedded in students' everyday study practices, from drafting essays to solving problems and preparing for exams, and is used both for legitimate learning support and for

* Correspondency: emrahbaki@beun.edu.tr

borderline or overt academic misconduct, intensifying concern about plagiarism, ghost-writing, and the evidential value of coursework and exams (Abdallah et al., 2025; Dhakal, 2025; Dos, 2025; Munaye et al., 2025). GenAI also produces fluent, human-like text that can be difficult to detect even for experts, undermining traditional plagiarism-detection approaches and raising questions about how to demonstrate authentic student work (Giannakos et al., 2024; Weber-Wulff et al., 2023). Although educators worry that overreliance on AI could erode students' writing and critical thinking, GenAI can also be used productively as a tutoring aid or a source of personalized feedback, if bounded by clear ethical expectations (Bobula, 2024; Weng et al., 2024; Xia et al., 2024). Across disciplines, GenAI requires universities to revisit plagiarism policies and reconsider assignments and exam design (Giannakos et al., 2024; Overono & Ditta, 2023).

Universities initially responded to GenAI in disparate ways. Early commentaries describe outright bans that treated AI-generated work as cheating (Cotton et al., 2024; Overono & Ditta, 2023). These blanket bans were quickly criticized as unsustainable and misaligned with preparing AI-literate graduates (Chan, 2023; Evangelista, 2025). Prohibition-only approaches may drive AI use underground rather than encourage transparent, responsible engagement (Chan, 2023). More recent work instead describes a shift toward pragmatic policies that acknowledge GenAI's presence and seek to guide its responsible use in teaching, learning, and assessment (Dai et al., 2024; Deng & Joshi, 2024; Moorhouse et al., 2023; Ullah et al., 2024).

By mid-2023, many universities had begun issuing official GenAI guidelines in coursework and assessment (Dai et al., 2024; Moorhouse et al., 2023; Ullah et al., 2024). These policies vary widely in tone and detail. Some remain conservative, treating undisclosed AI use as plagiarism, while others allow integration under specified conditions (Atkinson-Toal & Guo, 2024; Moorhouse et al., 2023). What remains less consistent is the extent to which these policies translate broad principles into operational assessment guidance, such as assignment-level rules, verification procedures, or redesign support for instructors (Deng & Joshi, 2024; Overono & Ditta, 2023; Weng et al., 2024; Xia et al., 2024).

Prior cross-institutional reviews also reveal limitations in the existing literature. Ullah et al. (2024) find wide variation in scope and specificity of institutional guidelines: most articulated high-level principles such as integrity, fairness, and transparency, but fewer provided operational guidance on assignment design, verification, or discipline-specific practice. Moorhouse et al. (2023) similarly show that policies often distinguish between allowed and prohibited uses, yet rarely offer structured templates, checklists, or decision tools for assessment redesign. Dai et al. (2024) highlight cautious adoption patterns in Asian universities, while Atkinson-Toal and Guo (2024) note that most guidance is advisory and principle-based. This matters because policy documents may contain more operational detail than secondary reviews capture.

For example, Monash University requires Chief Examiners to specify where and how AI may be used in assessments, and ETH Zurich requires transparent declaration of GenAI use while linking acceptable use to supervised and orally examinable assessment contexts. These examples suggest that primary policy texts may reveal more concrete assessment-design signals than are visible in broad review-level summaries.

Assessment redesign in the GenAI era

Research on assessment redesign increasingly emphasizes tasks that make student reasoning and process more visible. Weng et al. (2024) document a shift toward assessments

emphasizing higher-order thinking, multimodality, and process evidence, often integrating oral, practical, and in-class components. Xia et al. (2024) similarly identify multi-stage tasks, authentic and workplace-aligned problems, oral defenses, and reflective components that surface students' reasoning, while Mao et al. (2024) argue for combining AI-supported and AI-restricted tasks, with explicit attention to transparency of AI use and student agency. At course level, case studies report modified home exams, oral follow-ups, low-stakes tasks, and AI-aware designs such as AI-supported drafting plus in-class verification, structured peer review using AI suggestions, and oral defenses of AI-assisted projects (Farazouli et al., 2023; Foung et al., 2024). However, these examples remain mostly local and discipline-specific and are rarely consolidated into cross-institutional, reusable assessment patterns. Evidence focused on student experience is also less represented in cross-institutional policy analyses than instructor-facing redesign guidance, leaving a gap between institutional rules and learners' experience of them.

Integrity, fairness, privacy, and data governance

Beyond integrity, ethical questions about fairness, privacy, and data governance have become increasingly prominent. Research examines how AI-mediated assessment intersects with equity, bias, and surveillance (García-López & Trujillo-Liñán, 2025; Weng et al., 2024). Requiring students to upload coursework, voice data, or personal information into commercial GenAI tools can raise serious privacy and consent issues, especially when vendors retain or reuse user content (Giannakos et al., 2024; Xue, 2025). Detection-oriented responses, such as automated classifiers for AI-generated text, have also been shown to be unreliable and to disproportionately misclassify texts from second-language writers (Weber-Wulff et al., 2023).

These concerns have shifted the field toward "AI-aware" rather than "AI-blind" assessment, where AI use is anticipated and structured rather than treated as covert misconduct (Overono & Ditta, 2023; Xia et al., 2024). Proposed responses include documenting AI use, incorporating live or oral verification components, and designing tasks that foreground personal experience, disciplinary reasoning, or contextualized decision-making in ways that are harder to outsource fully to GenAI (Evangelista, 2025; Mao et al., 2024). However, the literature more often offers conceptual principles or discipline-specific vignettes than policy-grounded, step-by-step examples. Moreover, although concerns around fairness, privacy, consent, retention, and accessibility are increasingly acknowledged, they are not always examined systematically in institutional assessment guidance.

Frameworks and design patterns for AI-aware assessment

A small number of frameworks now attempt to systematize GenAI integration into assessment. Perkins et al. (2024) introduce the Artificial Intelligence Assessment Scale (AIAS), which positions assessment tasks along a five-level continuum from "No AI" through various forms of AI-supported and AI-generated work. Furze et al. (2024) illustrate AIAS-informed redesign in practice, while Kılınç (2024) extends the scale through the Comprehensive AI Assessment Framework (CAIAF), adding explicit attention to ethics, educational level, and advanced GenAI capabilities. Together, these frameworks provide useful conceptual scaffolds for making decisions about the role of AI in assessment. However, their main purpose is to classify or guide assessment design from a framework perspective. They do not show how assessment-related design responses are already being articulated in university policies and faculty-facing guidance.

Existing policy reviews, ethical discussions, and assessment redesign frameworks therefore

offer three important but only partially connected insights. One strand describes how universities are responding to GenAI at policy level. A second provides conceptual models for classifying or redesigning assessment. A third offers local course-level examples of adapted tasks and verification practices. What remains less developed is a policy-grounded synthesis showing how these strands come together in public institutional guidance and what recurring assessment configurations are visible to educators in practice. This gap matters because universities are no longer only stating whether AI is allowed; they are increasingly signalling how assessment should be redesigned when AI is available.

Pattern-based approaches in educational design offer a complementary lens. Design patterns capture reusable solutions to recurring pedagogical problems in a form that is concrete enough for practitioners to apply but abstract enough to transfer across contexts (Goodyear & Yang, 2009). Later work extends this logic to educational technologies and learning games, showing how pattern libraries can bridge the gap between scattered examples and systematic design guidance (Chimalakonda & Nori, 2021). This lens is useful here because it supports identifying recurring, transferable solutions to assessment-related problems without reducing institutional guidance to abstract principles alone. At the same time, pattern-based approaches are not neutral templates to be adopted mechanically; their value depends on interpretation in relation to context, purpose, and local constraints. In the present study, pattern-based educational design provides the theoretical grounding for interpreting institutional guidance as traces of emerging assessment design logic rather than isolated policy statements.

The present study

In response to the gap, the present study examines official, public-facing GenAI guidance from a cross-regional sample of English-medium universities in Europe, North America, Australasia, and Asia. English-medium universities were selected because publicly accessible guidance in a shared language makes it possible to examine policy wording, operational detail, and assessment-related design signals across regions while treating the documents as directly comparable primary texts. The study focuses on how universities frame GenAI in coursework and assessment, how actionable and assessment-specific their guidance is. It also examines the recurring assessment design patterns visible across institutions and the extent to which ethical concerns such as fairness, privacy, consent, retention, and accessibility are addressed. In doing so, it moves beyond broad policy description toward a more practice-oriented synthesis of how AI-aware assessment is articulated at the institutional level.

This gap informs the research questions below:

(1) What are the dominant themes and directives in current university policies on GenAI use in coursework and assessment?
(2) To what extent do these policies offer clear, actionable, and assessment-specific guidance for instructors, and what concrete tools do they provide for adapting assessment design and procedures in response to GenAI?
(3) What recurring assessment design patterns are universities proposing to maintain academic integrity and learning quality in the era of GenAI?
(4) How do current guidelines address concerns around fairness, bias, privacy, and accessibility in AI-supported assessment?

Method

Research design

This study used a framework-based document analysis of institutional policies and guidance on GenAI in higher education coursework and assessment. The design combined (a) directed content analysis of official policy texts, (b) descriptive and semi-quantitative cross-case summaries, and (c) a rule-governed pattern-construction stage. It examined how universities frame GenAI-related assessment guidance, how actionable that guidance was for instructors, and which recurring configurations could be synthesized into provisional assessment design patterns. The unit of analysis consisted of institutional policy or guidance documents on GenAI in coursework and assessment, including clearly linked subpages that formed part of the same institutional guidance set.

Corpus and sampling strategy

The corpus consisted of official, public-facing institutional policies and teaching-unit guidance documents explicitly addressing GenAI in coursework and assessment. Sampling was purposive and criterion-based. The final corpus included universities from Europe (n = 6), North America (n = 6), Australasia (n = 5), and Asia (n = 4). The aim was not statistical representativeness but cross-context analytic coverage of public-facing assessment-related GenAI guidance. Institutions were selected through a criterion-based process: English-medium universities were prioritised when they had visible, centrally hosted guidance on GenAI in coursework and assessment. This excluded non-English-medium institutions and limited the corpus to globally visible English-language guidance.

Each institution's relevant teaching-and-learning and assessment pages were searched using terms such as "GenAI," "ChatGPT," "artificial intelligence," "AI in assessment," and "AI policy." Candidate documents were recorded in a Corpus Log with URLs, access dates, and inclusion decisions (Appendix B). Data were collected between August 2025 and November 2025, and all URLs were last accessed on 05.11.2025. As GenAI policy was evolving rapidly during this period, the most current visible version was used, and access dates for each source were recorded in Appendix B.

The goal was to assemble a corpus sufficient to support pattern construction for assessment redesign. Sampling was concluded when the corpus provided cross-regional coverage, included the main document types through which universities communicated GenAI guidance, and yielded sufficient recurrence in assessment-related configurations to support pattern comparison. The sampling endpoint was therefore based on analytic coverage and pattern sufficiency.

An overview of the final corpus is provided in Table 1.

Table 1. Corpus overview

Institution	Region	Country
ETH Zurich	Europe	Switzerland
KU Leuven	Europe	Belgium
University College London (UCL)	Europe	UK
University of Cambridge	Europe	UK
University of Glasgow	Europe	UK
University of Leeds	Europe	UK

Institution	Region	Country
Harvard University	North America	USA
Penn State University	North America	USA
University of California, Berkeley	North America	USA
University of California, Los Angeles (UCLA)	North America	USA
University of Michigan	North America	USA
University of Minnesota	North America	USA
Monash University	Australasia	Australia
University of Otago	Australasia	New Zealand
University of Queensland	Australasia	Australia
University of Sydney	Australasia	Australia
UNSW Sydney	Australasia	Australia
National University of Singapore	Asia	Singapore
University of Hong Kong	Asia	Hong Kong
Nanyang Technological University	Asia	Singapore
The Hong Kong University of Science and Technology	Asia	Hong Kong

Document identification and selection

Documents were eligible for inclusion if they were (a) issued by a university; (b) hosted on an official institutional website and freely accessible without login; (c) focused on GenAI in coursework and assessment; and (d) publicly current at the time of data collection. Documents were excluded if they (a) addressed research use only; (b) focused on non-university sectors; or (c) were opinion pieces, news articles, or blogs. Where multiple webpages formed part of a single guidance set, clearly linked sub-pages were treated as one institutional case.

Initial document identification, screening, and inclusion decisions were conducted by the primary researcher using the stated criteria. This may have introduced selection bias and was partly minimised through an auditable corpus log and explicit inclusion/exclusion rules.

The document-identification process is summarised in Table 2. Institutional website searches initially yielded 78 potentially relevant pages. After preliminary screening, 54 candidate pages were assessed for relevance to GenAI in coursework and assessment. Of these, 33 were excluded. The final corpus comprised 21 university cases represented across 26 source webpages.

Table 2. Document identification and selection flow

Stage	Description	n
1	University websites and policy/guidance pages initially identified through institutional website searches	78
2	Candidate pages/documents screened for relevance to GenAI in coursework and assessment	54
3	Excluded after screening because they focused only on research use, were non-authoritative/news items, were duplicates, or did not address coursework/assessment	33
6	Final institutional cases included in the corpus	21 universities
7	Total source pages/URLs represented across the final 21 cases	26 webpages

Coding framework and coding manual

A directed content analysis approach was used, informed by prior work on GenAI, ethical AI governance, and emerging frameworks for AI-aware assessment. The coding framework operationalised policy content across the following core dimensions: disclosure requirements; permitted and prohibited uses; detection and verification practices; assessment and redesign guidance; privacy and data-governance safeguards; fairness, bias, and accessibility; concrete examples or templates; and three summary ratings for clarity, actionability, and assessment specificity.

Each institutional case was rated on a 3-point scale (0–2) along three axes. The clarity rating reflected the specificity and internal consistency of statements about GenAI. Actionability indicated whether instructors were given concrete steps, examples, or assessment strategies they could implement. Assessment specificity referred to the extent to which guidance was explicitly tied to assessment tasks. A condensed version of the coding dimensions and decision rules is provided in Table 3, while the full coding manual is provided in Appendix A.

Table 3. Condensed coding dimensions and decision rules

Coding family	Core variables	Values / format	Main decision rule
Policy content	Permitted AI uses; prohibited AI uses; disclosure required	Short text; Yes/No + note	Only explicitly stated permissions, prohibitions, and disclosure requirements were coded
Detection and verification	Detection tools stance; verification step present	Encourage / Discourage / Cautious / Silent; Yes/No + type	Reflected the dominant position in the document. Verification was coded only when a concrete mechanism was specified.
Assessment design features	Concrete assessment examples/patterns; live/online variants; group work controls	Yes/No + quote; Yes/No/Partial; Yes/No + note	Coded only when the document provided task-specific or mode-specific guidance relevant to assessment design.
Ethical and governance safeguards	Public uploads of identifiable data disallowed; retention policy stated; consent language; bias/fairness checks; accessibility alternatives	Yes/No; Yes/No; Yes/No/Partial; Yes/No; Yes/No/Partial	Coded only when explicitly linked to GenAI use in coursework or assessment.
Summary ratings	Clarity; actionability; assessment specificity	0 / 1 / 2	Ratings captured the degree to which guidance was clear, practically usable for staff, and directly tied to assessment tasks.
Illustrative evidence	Exemplar quote	≤ 2 sentences or blank	A short verbatim extract was recorded when useful to preserve a representative policy statement for interpretation and pattern construction.

Coding procedures and reliability

Coding proceeded in three stages: piloting, full coding, and reliability checking. First, two researchers independently coded four documents selected to reflect different document types and institutional contexts. This pilot phase tested coding boundaries, resolved ambiguities, and refined scale descriptors and building-block flags. Second, the full corpus was coded by the primary researcher. Third, reliability was assessed using documents from five of 21 universities, coded independently by a second coder using the finalized framework. These five documents were selected to represent different regions and document types.

For the ordinal 0–2 scales, percent agreement and quadratic-weighted Cohen's kappa were examined. Percent agreement ranged from .80 to 1.00. Because of restricted variance in some codes, kappa estimates were unstable or could not be estimated; therefore, Gwet's AC1 was also reported as a more stable chance-corrected coefficient under prevalence/imbalance conditions (AC1 = .76–1.00). For categorical policy-feature flags, percent agreement and AC1 were calculated (agreement range: .00–1.00, M = .71; AC1 range: −.41 to 1.00, median = .76, M ≈ .61). Full code-level reliability statistics are presented in Appendix A (Table A1).

The primary coding difficulties concerned the coding of consent language and accessibility alternatives. For consent language, disagreements clustered around the boundary between Partial and No/Yes. For accessibility alternatives, initial disagreement was systematic, indicating that the rule required clarification. These cases were reviewed, discussed, and resolved through explicit decision rules, which were incorporated into the final coding manual. All disagreements were adjudicated through discussion, after which the framework was finalized and applied consistently across the final dataset.

Data analysis and pattern construction

Analysis proceeded in three linked steps: descriptive cross-case mapping, gap mapping and pattern construction. First, frequency distributions were calculated for the three 0–2 ratings and for assessment-relevant building-block features. Because the three summary ratings showed restricted variance, they were used only to indicate whether an institutional case showed a general movement from principle-level guidance toward more operational assessment guidance. They were therefore retained as a compact descriptive device, but not treated as fine-grained measures for comparing institutions. However, the ratings had limited ability to distinguish meaningfully among institutional cases under the present sample conditions. Therefore, the main analytic weight rests on coded policy features, assessment-relevant building blocks, recurring pattern configurations, and ethical gap mapping rather than on score-based comparison. Second, cross-tabulation of ethical/governance dimensions with assessment-operational features was used to identify cases where integrity principles were present but procedural assessment guidance was limited, relying especially on variables coded as "Yes," "No," or "Partial." variables. Third, pattern construction used was based on interpretive comparison of cases with stronger operational guidance, indicated by higher clarity, actionability, and assessment specificity ratings, together with multiple relevant building blocks. Recurrent combinations of (a) boundaries for AI-use, (b) verification procedures, (c) required process evidence, and (d) rubric/assessment adjustments were identified and normalized into a common schema.

A configuration was formalized as a candidate pattern when it appeared in at least three institutions and included sufficient procedural detail to specify permitted AI use, required process evidence, and a verification mechanism. To focus the pattern library on stronger cross-institutional regularities, only patterns appearing in at least 11 institutions, a majority of the university corpus, were reported in the Results section.

The steps were sequential and cumulative: Step 1 established the descriptive profile of the corpus; Step 2 identified ethical and operational gaps; and Step 3 compared recurring assessment configurations in cases with sufficient assessment-operational detail. Pattern construction proceeded from the coded matrix generated through Steps 1 and 2.

Trustworthiness

Trustworthiness was supported through transparent inclusion criteria, an auditable corpus log (Appendix B), structured coding rules (Appendix A), pilot-based refinement, independent second-coder checks, and explicit discrepancy adjudication. Because the study involved interpretive judgments about policy meaning and pattern recurrence, reflexive caution was maintained throughout analysis. The primary researcher analyzed the corpus from an assessment-design perspective and examined how institutional guidance could serve as a practical design resource. To reduce over-interpretation, coding was anchored in explicit textual evidence, candidate patterns were required to show recurrence across institutions, and discrepancy discussions with the second coder were used to clarify coding boundaries. Even so, the final pattern set remains an interpretive synthesis rather than a purely procedural output of the coding matrix.

Result

Policy directives on use, disclosure, and detection (RQ1)

The frequencies reported in this section are not mutually exclusive: a single institutional document could contain several permitted uses, several prohibited uses, and multiple procedural requirements. Only explicitly stated permissions, prohibitions, and disclosure requirements were coded; no permissions or prohibitons were inferred from general policy language.

Permitted and prohibited uses

Frequently permitted uses included brainstorming (14/21), outlining (13/21), drafting/redrafting with constraints (10/21), grammar/style support (12/21), and formative explanation/feedback support (11/21). Frequently prohibited uses included submitting largely unedited AI-generated work as one's own (18/21), using GenAI in explicitly AI-free exams/tests (12/21), and using AI to bypass required disciplinary work (6/21).

Seventeen universities also explicitly permitted staff-facing GenAI use for internal teaching tasks, typically with explicit caution to verify output quality. For example, Nanyang Technological University stated that students are allowed to use Gen AI in their assignments and that students are expected to declare any use of AI and explain how it is used. When using AI, students are ultimately responsible for the content generated. This wording reflects the broader logic of the corpus: AI use was generally accepted, but only under conditions of transparency and continued human accountability.

Disclosure and course-level rule setting

20/21 universities required explicit declaration of AI use, typically through an assignment statement, declaration statements, or course-level wording. One institution relied on general academic-integrity regulations without a specific GenAI disclosure rule. Three universities explicitly required each course to specify local AI-use conditions. Across the corpus, many institutions required instructors to state whether AI was disallowed, conditionally allowed, or intentionally integrated for specific tasks. UNSW Sydney, for example, stated that *"Students must be provided with clear instructions stating whether they can use ChatGPT or other forms of GenAI for each assessment or learning activity and if so, for what purpose."* Similarly, the Hong Kong University of Science and Technology noted that *"faculty members are required to communicate the chosen policy option to students in writing."* These examples

indicate that disclosure extended beyond student acknowledgment to instructor-defined rule setting.

Detection tools and integrity framing

16/21 universities explicitly discourage or limit AI-detection tools as primary evidence, two allow or encourage cautious use, one adopts a general "use with caution" stance, and two are silent. In the discouraging group, policies emphasize that detectors can be inaccurate, may disproportionately flag non-native writers, and should not be used as sole grounds for misconduct accusations. In institutions with restrictive detector language, policies typically require follow-up verification such as dialogue with students or process evidence. Penn State University, for example, noted that *"Faculty are discouraged from using AI text analysis software as sole evidence in academic integrity cases,"* while HKUST stated that *"solely relying on AI detection tools… is not the best solution… we recommend that you shift your focus to reviewing your assessment design."* Together, such statements suggest that the integrity framing in the corpus was shifting from technological detection toward assessment redesign and human verification.

Actionability and assessment-operational guidance (RQ2)

Most documents moved beyond generic integrity language and included practical guidance for adapting assessment in GenAI contexts. In practice, this guidance typically took the form of model syllabus language, disclosure formats, conditional-use distinctions, examples of alternative task design, or procedures for follow-up verification. A smaller subset acknowledged GenAI as an emerging issue but offered limited procedural guidance.

Three recurring features characterised the more operationally detailed documents. First, templates and checklists appeared in 17 of the 21 institutions, often in the form of model syllabus statements, AI-use declarations, or implementation prompts for staff. Second, assessment-susceptibility typologies appeared in six institutions, where the guidance differentiated among task types in terms of how vulnerable they were to unverified AI-generated substitution. Third, a smaller subset provided stepwise redesign guidance, helping staff audit existing tasks and decide where additional process evidence or verification requirements might be needed.

A similar pattern emerged in the distribution of assessment-relevant building blocks. Verification mechanisms appeared in 19/21 universities and included in-class micro-writing, short oral checks, spot-check interviews, and comparison against prior drafts or known student performance. Process evidence appeared in 16/21 universities and included draft histories, AI-use logs, reflective notes and documented explanations of how AI had been used. Guidance on constrained tool use also recurred: five universities recommended restricting use to institutionally supported platforms, 19 explicitly warned against entering personal, confidential, or identifiable data into public AI tools, and ten clearly identified contexts in which no AI use was allowed. Finally, seven universities explicitly linked AI-related expectations to rubric or criteria design, making clear that students remained accountable for errors introduced by AI.

This actionability was especially visible where documents moved from principle to operational instruction. UNSW Sydney, for instance, stated that *"Students need course-specific instructions laying out the extent to which they can use AI for assessments and learning activities."* Likewise, the University of Hong Kong observed that teachers *"must critically evaluate current assessment methods that may be susceptible to student use of GenAI and consider replacing*

them with alternatives." These examples indicate that actionability in the corpus was tied to redesign-oriented guidance.

Recurring assessment design patterns (RQ3)

The third research question asked what recurring assessment design patterns universities were promoting to maintain learning quality in contexts where students have access to GenAI. Four patterns met the predefined reporting threshold and are presented in Table 4.

Table 4. Recurring Assessment Design Patterns (RQ3)

Pattern	n (%)	Target assessment purpose	AI-use conditions	Verification steps	Required student evidence	Rubric / assessment implications
Process-evidence portfolios and AI-use logs	19 (90.5%)	Preserve evidentiary value through visible process	Bounded AI use with transparent documentation	Draft review, staged submission, reflection, follow-up dialogue	Draft histories, AI-use declarations, prompt/output logs, reflective notes	Greater weight on process, development, reflection, and responsible AI use
AI-permitted drafting plus in-class or oral verification	15 (71.4%)	Allow bounded AI support while protecting authorship	AI allowed for brainstorming/drafting under stated conditions	Oral defense, viva, in-class writing, spot checks	AI-use declaration and live demonstration of understanding	Evaluation extends beyond product to defended reasoning and understanding
Critical engagement with AI outputs	12 (57.1%)	Develop evaluative judgment through critique of AI output	AI output used openly for analysis and critique	Comparison, commentary, error identification, revision justification	Critique, corrected outputs, annotated comparisons, reasoned revisions	Greater weight on critical reasoning, judgment, and justified evaluation
Secure high-stakes exams plus AI-enabled coursework	11 (52.4%)	Balance evidentiary control with flexible coursework	AI restricted in secure tasks, conditionally allowed in open coursework	Invigilation, supervised writing, disclosure/process checks	Disclosure statements, process records where relevant, secure exam performance	Evidentiary burden distributed across task types

Pattern 1 – Process-evidence portfolios and AI-use logs

The most prevalent pattern, identified in 19 universities, shifted part of the assessment evidence from the final product to the documented process of task completion. Here, AI use was permitted under clearly defined conditions only when students provided evidence such as draft histories, AI-use logs, reflective notes, annotated records, or staged submissions. The underlying rationale was that process visibility could preserve authorship, support accountability, and reduce reliance on the final text alone. ETH Zurich, for example, required that *"for any assessment where AI is permitted, students must include a statement acknowledging AI use, naming the tool and version, publisher, URL, and briefly describing how it was used; coordinators may additionally require logs of prompts/outputs."*

Pattern 2 – AI-permitted drafting plus in-class or oral verification

The second pattern, identified in 15 universities, permitted AI-supported drafting or brainstorming under stated conditions while preserving evidentiary confidence through follow-up verification. Student authorship was verified through oral defense, brief viva, in-class writing, short verification interviews with selected students, or similar mechanisms. UNSW Sydney described "viva-style conversations on process and meaning of their work" and "viva voce / oral components to test authorship and understanding of artefacts." Similarly, HKUST recommended "reflection papers with an oral defense" and "written or oral synchronous assessments" when AI use was not allowed.

Pattern 3 – Critical engagement with AI outputs

The third pattern, identified in 12 universities, treated AI output as material for analysis, critique, correction, or extension rather than as an undisclosed production aid. Assessment evidence is centred on evaluative commentary, reasoned correction, justified revision, or comparison between AI-generated and student-generated work. Nanyang Technological University, for instance, described an assessment design in which teachers could "ask students to use ChatGPT to generate an answer, critique the chatbot's replies, and suggest how the answers can be improved."

Pattern 4 – Secure high-stakes exams plus AI-enabled coursework

The fourth pattern, identified in 11 universities, combined AI-restricted high-stakes assessment tasks with more flexible AI-enabled coursework. Secure, supervised, or invigilated tasks were maintained where strong evidentiary control was needed, while other coursework allowed conditional AI use under disclosure or process-documentation requirements. This reflects a portfolio-based assessment logic in which different tasks carry different evidentiary burdens. The University of Sydney expressed this distinction clearly: "From Semester 2, 2025, you're generally not permitted to use AI in secure (supervised) tasks… For open (unsupervised) assessments, you will be able to use AI, and need to appropriately acknowledge its use."

Ethical safeguards, fairness, privacy, and accessibility (RQ4)

The fourth research question examined how current guidelines address fairness, bias, privacy, and accessibility. Table 5 presents the ethical gap map. Two areas were consistently strong: transparency mechanisms (disclosure and verification) and public-upload caution.

Disclosure was required in 20/21 documents, verification steps appeared in 19/21, and public uploads were either disallowed or strongly discouraged in 19/21.

Coverage was more limited regarding data retention and consent language. Only 7/21 documents stated retention policy details, and 6/21 provided explicit consent wording (with 4/21 showing partial reference and 11/21 no explicit consent text).

All documents referenced bias and fairness risks (21/21), though procedural depth varied. Accessibility/accommodations were explicitly addressed in 15/21, partially addressed in 4/21, and absent in 2/21.

Table 5. Ethical Gap Map Across 21 Universities

Ethical dimension	Yes n (%)	Partial n (%)	No n (%)
Disclosure required	20 (95.2%)	0 (0.0%)	1 (4.8%)
Verification step present	19 (90.5%)	0 (0.0%)	2 (9.5%)
Public uploads disallowed / strongly discouraged	19 (90.5%)	0 (0.0%)	2 (9.5%)
Retention policy stated	7 (33.3%)	1 (4.8%)	13 (61.9%)
Consent language provided	6 (28.6%)	4 (19.0%)	11 (52.4%)
Bias / fairness risks mentioned	21 (100.0%)	0 (0.0%)	0 (0.0%)
Accessibility / accommodations addressed	15 (71.4%)	4 (19.0%)	2 (9.5%)

Note. "Yes" = explicit treatment of the issue, "Partial" = implicit or high-level reference without concrete wording or procedures, "No" = no identifiable reference in the coded institutional documents.

Caution regarding public AI uploads was often articulated through explicit procedural guidance. The University of Sydney instructed students to "never enter personal or sensitive information into AI tools," and HKUST similarly advised staff and students "not to post any data that raises privacy concerns when using free or public GenAI services." By contrast, consent language was less developed. Although Nanyang Technological University required that "written permission has been explicitly provided by the data owner" before certain data could be used in GenAI tools, such explicit procedural guidance was uncommon across the corpus.

Discussion

This study examined how 21 English-medium universities across regions address GenAI in coursework and assessment through institutional policies and teaching and learning guidance. The findings suggest a shift beyond blanket prohibition toward more structured, assessment-aware guidance, but also persistent gaps in data governance, consent, and support for equitable access. These conclusions should be interpreted in light of the corpus: a purposive sample limited to publicly visible English-medium guidance and skewed toward institutions with more developed public policy communication.

From prohibition to structured permission (RQ1)

The policy landscape observed in this study aligns with earlier reviews of institutional responses to GenAI. These reviews describe an initial phase of uncertainty and sporadic bans, followed by more pragmatic and regulative approaches (Chan, 2023; Dai et al., 2024; Deng & Joshi, 2024; Moorhouse et al., 2023; Ullah et al., 2024). All institutions in the corpus now allow some student use under conditions typically distinguishing between learning-support uses such as brainstorming and higher-risk uses such as submitting unedited AI-generated work. This pattern mirrors the "benefit versus misconduct" tension highlighted in broader reviews (Abdallah et al., 2025; Dhakal, 2025; Dos, 2025; Munaye et al., 2025; Weng et al., 2024; Xia et al., 2024) and in integrity-focused commentaries warning against both uncritical uptake and blanket rejection (Cotton et al., 2024; Overono & Ditta, 2023).

The widespread requirement for disclosure suggests that transparency around AI use has become a core integrity expectation. Rather than treating any AI involvement as cheating, policies require students to declare AI use and emphasize continued responsibility for the accuracy and integrity of AI-assisted work. This aligns with calls for "AI-aware" rather than "AI-blind" assessment, where AI use is anticipated and structured rather than treated as covert misconduct (Overono & Ditta, 2023; Xia et al., 2024). That all sampled institutions allow some student use of GenAI may reflect both pedagogical adaptation and the practical difficulty of enforcing blanket prohibition. The present data cannot fully distinguish between these interpretations.

A second notable shift concerns AI-detection tools. Most universities explicitly discourage or limit reliance on detectors and caution against using them as sole evidence for misconduct. This reflects the documented unreliability of current detectors and their tendency to misclassify second-language writers (Weber-Wulff et al., 2023), as well as broader concerns about fairness and false accusations (García-López & Trujillo-Liñán, 2025; Weng et al., 2024). Instead, policies recommend conversational checks, comparisons with prior work, and process evidence, which are more labor-intensive but better aligned with both integrity and due process. The corpus thus reflects a position closer to critics of detection-heavy approaches, who argue that retrospective policing is less sustainable than proactive assessment redesign (Chiu, 2023; Cotton et al., 2024; Xia et al., 2024).

Beyond principles: clarity, actionability, and assessment specificity (RQ2)

Earlier reviews report that many institutional GenAI policies remain high-level, emphasizing integrity, fairness, and transparency while offering limited operational guidance for assessment (Atkinson-Toal & Guo, 2024; Dai et al., 2024; Moorhouse et al., 2023; Ullah et al., 2024). In the present corpus, the key finding was not simply that documents scored highly on summary ratings, but that many of them included concrete assessment-operational features that staff could use in practice. These included checklists, model syllabus language, AI-use notes, verification procedures, process-evidence requirements, constrained-use guidance, and, in a smaller subset, task-susceptibility typologies and stepwise redesign prompts. This pattern is more consistent with structured support for decision-making than with the abstract, warning-style guidance described in earlier reviews (Perkins et al., 2024; Weng et al., 2024; Xia et al., 2024).

Taken together, these features suggest that, at least in this sample, guidance is becoming more operationally specified. Compared with the "principle-heavy, example-light" guidance described by Moorhouse et al. (2023) and Ullah et al. (2024), the policies in this corpus already

embed concrete building blocks for redesign, including verification mechanisms, process evidence requirements, constraints on tool use, and rubric-level adjustments. Although implementation may still vary by disciplinary context, this marks a shift from merely signalling that "GenAI is an issue" toward supporting assessment redesign.

At the same time, this interpretation should not rest too heavily on the three 0–2 summary ratings of clarity, actionability, and assessment specificity. Because most documents clustered at the top end of the scale, the ratings functioned as broad indicators of general operational orientation rather than fine-grained comparative measures. The sample also favored institutions with relatively mature and publicly visible guidance. For that reason, the greater interpretive weight is therefore placed here on concrete features, excerpts, and pattern configurations than on summary scores alone.

The distribution of those features was also uneven. Verification mechanisms and process evidence were near-ubiquitous, but explicit linkage to rubric criteria appeared in only a minority of policies, and AI-specific guidance for group work was rare. This reflects broader discussions in the literature: although authors frequently recommend rubrics that value reasoning, reflection, and AI literacy (Evangelista, 2025; Mao et al., 2024; Weng et al., 2024; Xia et al., 2024), the documents in this sample focus more on task configurations and procedural safeguards than on how learning should be evidenced or how assessment criteria should be adjusted.

Institutional design patterns and existing frameworks (RQ3)

By synthesising recurrent combinations of building blocks, the study contributes a provisional policy-grounded set of assessment design patterns recommended in response to GenAI. The four dominant patterns resonate with strategies proposed in conceptual work and case studies: process-evidence portfolios and AI-use logs, AI-permitted drafting plus in-class or oral verification, critical engagement with AI outputs, and secure high-stakes exams plus AI-enabled coursework. Their distinctive contribution, however, is that they are derived from cross-institutional policy texts rather than from individual course-level examples. Together, they show that responses to GenAI extend beyond simple bans or generic warnings to systematic combinations of process evidence, verification, constrained tool use, and diversified portfolios.

These patterns closely match promising GenAI-responsive assessment principles identified in recent reviews, including multi-stage tasks, process-based grading, oral or practical components, and combinations of AI-restricted and AI-enabled evidence (Chiu, 2023; Mao et al., 2024; Weng et al., 2024; Xia et al., 2024). Case studies of "GenAI-empowered assessment" similarly highlight AI-supported drafting plus in-class verification, structured peer review using AI suggestions, and oral defenses of AI-assisted work (Farazouli et al., 2023; Foung et al., 2024). The findings add cross-institutional evidence that such designs are reflected across multiple policy and guidance documents, rather than remaining isolated local innovations.

The patterns also provide a bottom-up complement to frameworks such as the Artificial Intelligence Assessment Scale (AIAS; Perkins et al., 2024) and the Comprehensive AI Assessment Framework (CAIAF; Kılınç, 2024). Whereas AIAS and CAIAF classify tasks and risks, the patterns identified here show how such classifications appear in course documentation and assessment briefs (Furze et al., 2024; Perkins et al., 2024). The findings suggest that this pattern library may help make these frameworks more usable in institutional settings, but its value should be understood as a practice-oriented proposition rather than a fully validated

implementation tool.

The pattern library helps bridge the gap between high-level frameworks and the concrete artifacts instructors must produce and aligns with work on pattern-based educational design valuing reusable, adaptable design knowledge (Chimalakonda & Nori, 2021; Goodyear & Yang, 2009). At the same time, patterns can be decontextualized, copied mechanically, or detached from the disciplinary, institutional, and resource conditions that make them workable. The present pattern set should therefore be read as recurring design logics visible in the sampled guidance, not as a universal recipe or a validated consensus across higher education. Its provisional status reflects that it was constructed from a small, non-representative corpus of institutions with publicly visible policy guidance rather than from direct observation of implementation across contexts.

Ethical safeguards: strengths and blind spots (RQ4)

The ethical gap map shows a mixed picture. The corpus shows stronger coverage across several dimensions highlighted in recent reviews. Nearly all institutions warn students not to upload personal or confidential data to public GenAI tools, reflecting concerns about vendor retention and secondary use of student content (García-López & Trujillo-Liñán, 2025; Giannakos et al., 2024; Xue, 2025). All policies mention bias and fairness risks in some form, and some explicitly caution that detectors may disproportionately flag certain student groups, echoing evidence that AI-generated-text detectors misclassify texts by second-language writers at higher rates (Weber-Wulff et al., 2023). Many documents also ask instructors to consider accessibility, for example, by checking that required tools meet institutional accessibility standards and that restrictions do not disadvantage students who rely on assistive technologies.

By contrast, retention and consent emerge as clear blind spots. Only a minority of policies state concrete retention periods or refer to specific vendor policies, and more than half provide no ready-to-use consent wording even where data sharing or recording is envisaged. In practice, instructors may be warned about data risk but left to craft their own consent documents, risking inconsistent or legally fragile practice. These gaps also raise regulatory concerns. In General Data Protection Regulation (GDPR)-governed contexts, unclear consent processes, vague retention provisions, or weak guidance on third-party data handling can create uncertainty about lawful processing, transparency, and minimization. In US contexts, similar issues may intersect with Family Educational Rights and Privacy Act (FERPA)-related concerns where student records or coursework are handled through external systems. The corpus therefore suggests that institutions are now relatively strong at identifying GenAI-related risks, but less developed in translating that awareness into governance language suitable for compliance-sensitive settings..

A similar pattern appears for fairness: awareness is widespread, but operational guidance remains thin. Policies commonly list generic risks, such as hallucinations, biased training data, and overreliance on detectors, but only a few offer concrete steps to mitigate inequity in assessment, such as providing non-AI alternatives where access is constrained or considering the implications of AI bans for disabled students. Overall, the gap map suggests that institutional policies have prioritised integrity and instructional guidance over data governance and consent and will need to evolve further as GenAI tools become more embedded in assessment systems.

Implications for policy and practice

For policymakers and teaching-and-learning units, the findings suggest several practical implications. First, the fact that many institutions already provide assessment-focused examples, templates, and patterns shows that it is feasible to move beyond high-level principles. Universities still at an early stage may benefit from using pattern-based representations, similar to those used in educational design and learning-game research, to capture and share their own emerging practice (Chimalakonda & Nori, 2021; Goodyear & Yang, 2009). These can also serve as starting points for local adaptation and refinement.

Second, the building blocks highlighted in this study provide a useful vocabulary for institutional conversations about GenAI. Assessment teams can use these building blocks, and the patterns that combine them, as a menu of options when redesigning programs, for example, deciding where to emphasize process-evidence portfolios, where to use AI-permitted drafting plus oral verification, and where to retain secure AI-free examinations. Because these elements are often scattered across policy documents, synthesising them into accessible guidance for program-level planning could be a valuable next step. At the same time, such redesign choices have labor and resource implications. Oral verification, iterative draft review, and process-based assessment increase marking time, require staff calibration, and may be more feasible in some class sizes and disciplines than in others. Practical adoption, therefore, depends not only on design logic but also on workload, staffing, and institutional support.

Third, the ethical gap map suggests immediate areas for improvement. Institutions could strengthen their policies by providing clearer, task-specific consent templates, specifying retention periods and linking to vendor policies where relevant, and offering more concrete advice on designing assessments that do not disadvantage students with disabilities or limited access to AI tools. These steps would address concerns about privacy, equity, and surveillance, while still supporting the assessment innovation recent reviews see as necessary (Chiu, 2023; Mao et al., 2024; Weng et al., 2024; Xia et al., 2024).

Implications also extend to students, as primary participants in AI-aware assessment. Clearer disclosure expectations, transparent task rules, and more explicit guidance on acceptable support uses may reduce uncertainty and perceived arbitrariness. Stronger accessibility provisions, non-AI alternatives where appropriate, and clearer consent language may also help protect students from disadvantage when AI access, digital confidence, or disability-related needs are unevenly distributed.

Finally, the practical reach of these implications is uneven across contexts. Institutions in non-English-medium or under-resourced settings may lack the capacity to produce detailed public guidance, maintain institutionally licensed AI tools, or implement labour-intensive verification procedures. The pattern set offered here may still be useful, but it should be adapted to local language, infrastructure, and resourcing constraints rather than transferred uncritically.

Limitations and Future Research

Several limitations should be considered when interpreting the findings. First, the study was based on a purposive corpus of publicly available, English-medium institutional policies and guidance documents. This supported cross-case comparison of visible policy responses, but did not capture internal documents, local disciplinary adaptations, or non-public institutional practices. It also excluded major higher education systems where guidance is not publicly available in English. The findings should therefore not be read as globally representative.

Second, the corpus likely overrepresents institutions with more developed central teaching and learning infrastructures and more visible web-based policy communication. This may partly explain the high scores on clarity, actionability, and assessment specificity. The near-ceiling effect on the three 0–2 ratings is therefore a substantive limitation: the scales were more useful for signaling broad operational orientation than for distinguishing meaningfully among stronger documents.

Third, the full corpus was coded by the primary researcher, with a second coder independently checking five of the 21 universities. Although explicit coding rules, double-coding, and discrepancy resolution strengthened consistency, researcher judgment still shaped both full-corpus coding and pattern interpretation. Pattern construction likewise involved interpretive synthesis rather than purely mechanical aggregation.

Fourth, the study captures a rapidly evolving policy context. The documents were accessed between August and November 2025, and some guidance may already have changed. The corpus should therefore be treated as a time-bound snapshot rather than a stable endpoint.

These limitations point directly to future research needs. Longitudinal studies are needed to track how institutional GenAI guidance evolves, especially as universities refine disclosure rules, consent language, and data-governance provisions. Research is also needed on enactment: policy documents indicate intended directions, but not how consistently assessment patterns are implemented or how students experience them. Classroom- and program-level studies could examine whether the patterns identified here actually support learning, integrity, and inclusion when enacted.

Future work should attend more directly to student perspectives, particularly because protections around consent, retention, and accessibility appeared less developed than integrity-related controls. Interview, survey, and ethnographic research could examine how students interpret AI-use rules, how confident they feel about disclosure, and whether AI-aware assessment is experienced as supportive, exclusionary, or procedurally unclear. Comparative research should also extend beyond English-medium and publicly visible policies through translation, institutional partnerships, or access to internal policy archives. Finally, further work is needed on policy design for ethical safeguards, especially actionable consent templates, retention policies, and accessibility guidance that are pedagogically usable, legally robust, and sensitive to local resource conditions.

Conclusion

Taken together, the findings address the four research questions in a coherent way. Current institutional GenAI guidance in this corpus emphasizes conditional permission, disclosure, and verification rather than blanket prohibition. It often provides concrete operational assessment guidance rather than only abstract principles. It also identifies four recurring assessment design patterns centred on process evidence, verification, critique, and diversified portfolios. Finally, it addresses fairness, privacy, and bias more consistently than consent, retention, and accessibility. The study therefore suggests that institutional GenAI guidance can be read not only as regulation but also as an emerging assessment design resource. At the same time, these resources remain uneven and provisional, shaped by the limits of a small, non-representative corpus. The study's contribution lies not in offering a definitive model of AI-aware assessment, but in making visible recurring institutional responses that can be tested, critiqued, and adapted across contexts.

Declarations

Acknowledgments: Not applicable
Funding: No funding was received.
Ethics Statements: This study analyzed publicly available institutional policy and guidance documents on generative AI use in higher education coursework and assessment. It involved no human participants, no personal or sensitive individual-level data, and no intervention.
Conflict of Interest: The author has no relevant financial or non-financial interests to disclose.
Informed Consent: Not required because the study did not involve human participants or identifiable personal data.
Data availability: Data supporting the findings (Appendices)are available on Figshare https://doi.org/10.6084/m9.figshare.31112704.v1

References

Abdallah, N., Katmah, R., Khalaf, K. & Jelinek, H. F. (2025). Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration. *Social Sciences & Humanities Open*. 12, 1-14. https://doi.org/10.1016/j.ssaho.2025.101866

Atkinson-Toal, A., & Guo, C. (2024). Generative artificial intelligence education policies of UK universities. *Enhancing Teaching and Learning in Higher Education, 2*(1), 70–94. http://doi.org/10.62512/etlhe.20

Bobula, M. (2024). Generative artificial intelligence in higher education: A comprehensive review of challenges, opportunities, and implications. *Journal of Learning Development in Higher Education, 30, 1-27*. https://doi.org/10.47408/jldhe.vi30.1137

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*(1), Article 38, 1-25. https://doi.org/10.1186/s41239-023-00408-3

Chimalakonda, S., & Nori, K. V. (2021). A patterns based approach for the design of educational technologies. *Interactive Learning Environments, 31*(4), 2114–2133. https://doi.org/10.1080/10494820.2021.1875000

Chiu, T. K. F. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments*, *32*(10), 6187–6203. https://doi.org/10.1080/10494820.2023.2253861

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International,* 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Dai, Y., Lai, S., Lim, C. P., & Liu, A. (2024). University policies on generative AI in Asia: Promising practices, gaps, and future directions. *Journal of Asian Public Policy*. 18(2), 260-281. https://doi.org/10.1080/17516234.2024.2379070

Deng, X. N., & Joshi, K. D. (2024). Promoting ethical use of generative AI in education. *ACM SIGMIS Database,* 55(3), 6–11. https://doi.org/10.1145/3685235.3685237

Dhakal, S. P. (2025). A scoping review of generative artificial intelligence and pedagogy nexus: Implications for the higher education sector. *Metrics, 2*(3), 11-15. https://doi.org/10.3390/metrics2030017

Dos, İ. (2025). A systematic review of research on ChatGPT in higher education. *The European Educational Researcher,* 8(2), 59 76. https://doi.org/10.31757/euer.824

Evangelista, E. D. L. (2025). Ensuring academic integrity in the age of ChatGPT: Rethinking exam design, assessment strategies, and ethical AI policies in higher education. *Contemporary Educational Technology, 17*(1), 1-19. https://doi.org/10.30935/cedtech/15775

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, *49*(3), 363–375. https://doi.org/10.1080/02602938.2023.2241676

Foung, D., Lin, L., & Chen, J. (2024). Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices. *Computers and Education: Artificial Intelligence, 6*, 1-7. https://doi.org/10.1016/j.caeai.2024.100250

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, *40*(4), 38–55. https://doi.org/10.14742/ajet.9434

García-López, I. M., & Trujillo-Liñán, L. (2025). Ethical and regulatory challenges of generative AI in education: A systematic review. *Frontiers in Education, 10*, 1-13. https://doi.org/10.3389/feduc.2025.1565938

Giannakos, M.N., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y.A., Leo, D.H., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. Behaviour & Information Technology, 44, 2518 - 2544. https://doi.org/10.1080/0144929X.2024.2394886

Goodyear, P. & Yang, D. F. (2009). Patterns and pattern languages in educational design. In L. Lockyer, S. Bennett, S. Agostinho, & B. Harper (Eds.), *Handbook of Research on Learning Design and Learning Objects: Issues, Applications, and Technologies* (pp. 167-187). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-59904-861-1.ch007

Kılınç, S. (2024). Comprehensive AI assessment framework: Enhancing educational evaluation with ethical AI integration. *Journal of Educational Technology and Online Learning, 7*(4), 521–540. https://doi.org/10.31681/jetol.1492695

Mao, J., Chen, B. & Liu, J.C. (2024). Generative artificial intelligence in education and its implications for assessment. *TechTrends* 68, 58–66. https://doi.org/10.1007/s11528-023-00911-4

Moorhouse, B. L., Yeo, M. A., & Wan, Y. W. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open, 5(15)*, 1-10. https://doi.org/10.1016/j.caeo.2023.100151

Munaye, Y. Y., Endalamaw, M. T., & Taye, M. M. (2025). ChatGPT in education: A systematic review on opportunities, challenges, and future directions. *Algorithms, 18*(6), 1-28. https://doi.org/10.3390/a18060352

Overono, A. L., & Ditta, A. S. (2023). The rise of artificial intelligence: A clarion call for higher education to redefine learning and reimagine assessment. *College Teaching*, *73*(2), 123–126. https://doi.org/10.1080/87567555.2023.2233653

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice, 21*(6), 1-18. https://doi.org/10.53761/q3azde36

Ullah, M., Bin Naeem, S., & Kamel Boulos, M. N. (2024). Assessing the guidelines on the use of generative artificial intelligence tools in universities: A survey of the world's top 50 universities. *Big Data and Cognitive Computing*, 8(12), 1-16. https://doi.org/10.3390/bdcc8120194

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltynek, T., Guerrero-Dib, J., Popoola, O., Sigut, P. & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal of Educational Integrity* ,19, Article 26, 1-39. https://doi.org/10.1007/s40979-023-00146-z

Weng, X., Xia, Q., Gu, M., Rajaram, K., & Chiu, T. K. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. *Australasian Journal of Educational Technology*, *40*(6), 37–55. https://doi.org/10.14742/ajet.9540

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21, Article 40, 1-22. https://doi.org/10.1186/s41239-024-00468-z

Xue, Y., Chinapah, V., & Zhu, C. (2025). A comparative analysis of AI privacy concerns in higher education: News coverage in China and Western countries. *Education Sciences, 15*(6), 1-25. https://doi.org/10.3390/educsci15060650