Article

Journal of Interdisciplinary Studies in Education

Volume 15, Issue 6 (2026), pp. 329-356
Journal of Interdisciplinary Studies in Education
ISSN: 2166-2681Print 2690-0408 Online | https://ojed.org/jise
https://doi.org/10.32674/n7pdhe28

Transforming Assessment in Higher Education in the Era of Generative AI: From Knowledge Control to Critical Thinking Development

Yuliia Zubtsova, Larysa Sushchenko, Liudmyla Shulha,
Olena Andryushchenko, Oleksandr Shulha
Zaporizhzhia National University, Ukraine

ABSTRACT

The wider use of generative AI has raised concerns about how student learning is assessed in higher education. Tasks once considered reliable no longer guarantee independent work, so assessment is shifting in focus. While final written outputs still matter, greater emphasis is now placed on how students engage with material, construct arguments, verify information, and demonstrate reasoning. Attention is also given to how digital tools are used, especially when they replace rather than support thinking. This article examines changes in assessment practices in response to generative AI, identifies vulnerable formats, and highlights those that better reflect genuine learning, based on research, policy, and university practice.

Keywords: assessment, academic integrity, critical thinking, generative AI, higher education

INTRODUCTION

Generative artificial intelligence (AI) has entered higher education not merely as another digital tool but as a force reshaping assessment practices. What was once considered a reliable way to evaluate knowledge is no longer fully certain. Homework essays, analytical papers, and individual written assignments can no longer automatically be treated as unquestionable evidence that a student independently engaged with the material and mastered the required skills. The Organisation for Economic Co-operation and Development (OECD) (2026) notes that generative AI is already influencing everyday learning and the functioning of schools, while warning that excessive reliance on systems providing ready-made answers may reduce deep engagement and weaken learning outcomes.

This challenge is especially visible in assessment formats that depend primarily on written production. When texts can be generated quickly, organised coherently, and presented persuasively, a fundamental question emerges: is assessment measuring students' knowledge or merely the quality of a generated response? Xia et al. (2024) argue that the reliability of traditional assessment methods is increasingly weakened because systems such as ChatGPT are already being used to produce essays, proposals, and take-home examination responses. As a result, even suspicion of AI involvement begins to alter assessment logic and evaluative judgement. Consequently, the issue extends beyond isolated cases of academic misconduct and reflects a broader structural transformation in higher education.

However, reducing the discussion solely to academic dishonesty would be overly simplistic. Generative AI has also exposed a long-standing weakness in university assessment: the strong emphasis placed on the final written product rather than the learning process itself. Assessment has often prioritised outcomes while paying less attention to how students develop arguments, manage uncertainty, revise ideas, and refine their thinking. The OECD report suggests that output-oriented assessment models become less effective when high-quality products can be generated without full intellectual engagement. Therefore, increasing importance is now attached to forms of evaluation that make students' reasoning and learning processes more visible.

This broader transformation is closely connected to debates in AI ethics and educational theory. Floridi et al. (2018) argue that AI systems should be understood not only as technical instruments but also as socio-ethical actors that reshape human decision-making, responsibility, and institutional trust. Within higher education, this perspective highlights that assessment is no longer solely a pedagogical issue but also an ethical one involving transparency, accountability, and fairness in human–AI interaction. At the same time, Luckin (2018) emphasises that AI in education should support "human intelligence amplification" rather than replace intellectual effort. From this perspective, assessment systems must increasingly evaluate how students critically engage with AI-supported learning processes rather than merely assessing polished final outputs.

In this context, the importance of metacompetences becomes particularly evident. The central learning outcome is no longer the mechanical reproduction of information. Since AI tools can rapidly retrieve material, propose structures, and improve stylistic quality, greater value is now placed on evaluating, interpreting, questioning, and justifying information. Xia et al. (2024) connect these developments with self-regulated learning, academic integrity, and the need to strengthen AI and digital literacy among both staff and students. Similarly, Zhan et al. (2025) emphasise that meaningful engagement with AI-generated feedback requires metacognitive reflection, ethical judgement, and critical evaluation rather than purely technical competence. These arguments also align with Luckin's (2018) view that future-oriented education must prioritise adaptability, critical thinking, and reflective learning in AI-rich environments.

At the same time, the risks associated with academic misconduct have not disappeared but instead evolved into new forms. Abbas et al. (2024) found that high workload and time pressure significantly increase students' reliance on ChatGPT, while excessive dependence may contribute to procrastination, weaker memory retention, and lower academic performance. Yusuf et al. (2024), drawing on data from participants across 76 countries, likewise report widespread awareness and willingness to use generative AI tools, alongside concerns regarding overreliance, academic dishonesty, and the absence of clear ethical guidelines. Universities therefore face a significant contradiction: they cannot ignore emerging technologies, yet traditional control mechanisms are becoming increasingly insufficient. Floridi et al. (2018) similarly warn that governance structures frequently lag behind technological change, creating uncertainty regarding responsibility, ethical regulation, and institutional adaptation.

Despite the rapid growth of research on generative AI in higher education, the field remains fragmented. Some studies focus primarily on risks, others emphasise opportunities, while still others examine institutional regulation. Far fewer attempts integrate these perspectives into a unified conceptual framework. Xia et al. (2024) note that understanding how generative AI is reshaping assessment in higher education remains emergent and theoretically inconsistent. This gap becomes particularly visible in discussions of assessment as a transition from simple knowledge verification toward the development of critical thinking in environments where AI is integrated into everyday learning. Integrating ethical perspectives from Floridi et al. (2018) and educational perspectives from Luckin (2018) strengthens this discussion by framing assessment transformation not only as a technical or institutional challenge but also as a broader redefinition of learning, responsibility, and intellectual development in the age of AI.

The novelty of this study lies in approaching the transformation of assessment not through a single isolated problem, but through the combination of four interrelated dimensions: academic integrity, institutional responses to the spread of generative AI, critical thinking as the main objective of assessment, and contemporary methods of evaluation. In this way, evaluation is understood not just as a control mechanism but also as a site where judgement, accountability, and intellectual independence are developed. Unlike existing guidance that treats these dimensions separately—sector frameworks such as QAA (2023a) and TEQSA (2023), which set out broad principles; Corbin et al. (2025), who frame AI and assessment as a context-dependent "wicked problem" rather than a design solution; and Xia et al. (2024), who map the research landscape through a scoping review—this study integrates the four dimensions into a single operationalisable framework with explicit, observable indicators for assessment design (Section 4.4).

In addition to examining how generative AI influences assessment in higher education, this study aims to support models that promote academic integrity and critical thinking. It addresses how AI changes standard assessment, which formats are more resistant, how it can be integrated without compromising integrity, and which strategies best support critical thinking. Methodologically, the study uses a review-based analytical design. It systematises current international literature, interprets findings, and develops a conceptual framework for rethinking university assessment.

LITERATURE REVIEW

The Evolution of Assessment in Higher Education: From Knowledge Control to Judgement Development

To understand why the emergence of generative AI has had such a strong impact on assessment, it is useful to briefly consider how assessment has evolved over time. For a long period, higher education relied primarily on a summative logic in which the student was expected to demonstrate a result, while the lecturer assessed how well that result met established requirements. This model worked effectively in situations where it was necessary to capture the level of knowledge acquisition, but it was far less effective in showing how students think, how they respond to feedback, how they revise their own mistakes, and whether they are able to evaluate the quality of their own work.

For this reason, recent research places increasing emphasis not so much on assessment as the assignment of a grade, but on assessment as an integral part of the learning process. In the work of Henderson et al. (2025), feedback research has clearly shifted from just reporting faults to focusing on how students interpret this information and apply it to future learning. According to this viewpoint, feedback is no longer just an adjunct to evaluation. Without it, the process of progress becomes shallow and turns into a form of learning activity.

This change contributes to the explanation of formative evaluation's increasing significance. Because it monitors learning as it happens rather than waiting for the outcome, it is beneficial. Before finishing the project, students have the chance to review their work and make any necessary changes. According to Usher (2025), this reasoning separates formative evaluation from summative assessment. While the latter concentrates on documenting the level previously attained, the former aims to encourage learning. Actually, this suggests that the emphasis moves from a single evaluation to the consistent gathering of information showing pupils' progress in their work.

In this case, the problem of student participation in feedback becomes particularly important. The mere fact that comments are received does not guarantee their effectiveness. Zhan and Yan (2025) demonstrate that a major issue in higher education is low student engagement with feedback rather than a general absence of feedback. If students do not evaluate the advice, apply it to their own work, and utilise it to direct their next actions, even clear, helpful, and well-targeted guidance will not have much of an impact. As a result, contemporary evaluation techniques are increasingly linked to the concept of feedback literacy, which is described as the ability to engage with feedback in a meaningful rather than merely formal manner.

This also highlights another crucial element, which is the growth of evaluative judgment. It refers to the ability not simply to wait for external evaluation, but to recognise the strengths and weaknesses of one's own work. Zhan and Yan (2025) emphasise that effective engagement with feedback requires more than technical or language skills. It also involves evaluative judgement, metacognitive control, emotional reflection, and ethical decision-making. This changes the very function of assessment. It is no longer limited to ranking or selection, but begins to operate as a tool for the student's intellectual development.

The role of peer assessment and other dialogic formats is also changing. Their value lies not only in providing more comments or reducing the lecturer's workload. Usher and Faraon (2025) show that, when properly organised, peer assessment supports the development of self-regulation, encourages more attentive engagement with assessment criteria, and promotes deeper involvement in the learning material. In other words, when students evaluate the work of others, they also learn to view their own work differently. This reflects one of the key shifts in contemporary assessment, as it begins to shape not only the outcome, but also the learner as an active subject of the educational process.

Thus, even before the rapid spread of generative AI, it had already become clear that the traditional model of knowledge control in higher education was gradually losing its dominant role. It was being replaced by a different logic, more focused on process, dialogue, and learner autonomy, and more attentive to judgement and reflection. Generative AI entered this already evolving, yet still unsettled, landscape (Ardito, 2025).

Generative AI as a Challenge to Traditional Assessment

The development of generative AI did not start this transformation. But it has undoubtedly expedited it and highlighted the shortcomings of conventional evaluation forms, particularly those that primarily depend on the final written result. According to Kehler et al. (2025) and Xia et al. (2024), students are already using technologies like ChatGPT to finish essays, proposals, and take-home tests. As a result, traditional methods of assessment are losing their ability to demonstrate what students have truly learned. This problem extends beyond instances of academic dishonesty. The question of whether a submitted text may still be regarded as trustworthy proof of a student's original thought is also brought up.

Experimental evidence supports this conclusion. According to Curtis (2025) and Kofinas et al. (2025), instructors are typically unable to accurately discern between work generated with generative AI and work finished without it. However, the appraisal process is already influenced by the suspicion of such use. The authors also contend that academic integrity is not adequately protected by authentic assessment alone. Higher education institutions cannot use this strategy as a general solution because of this.

Rethinking assessment design itself is one of the main points of contention in the current debate, rather than identifying wrongdoing. According to Kofinas et al. (2025), generative AI's expanding capabilities undermine outcome-oriented written formats and promote a move toward more interactive, activity-based, and interpersonal modes of assessment. The crucial answer in this situation is not technological detection, but rather creating assignments that allow for the evaluation of not just the text but also the student's ability to defend their argument, think critically, and make proper conclusions.

However, current research does not view AI intelligence as a threat in and of itself. According to Henderson et al. (2025) and Wang et al. (2025), students appreciate the accessibility, speed, volume, and clarity of AI-generated feedback. However, the same studies also demonstrate that instructor feedback is thought to be significantly more trustworthy, particularly in circumstances where trust, discipline knowledge, and contextual awareness are crucial. Based on this, the authors conclude that AI-generated feedback and lecturer feedback are not interchangeable, but rather complementary.

Usher (2025), who compares evaluation and feedback generated by chatbots, students, and a lecturer, reports a similar outcome. According to the study, chatbots frequently give more thorough advice and typically offer higher scores than people. However, peer evaluation more accurately captures contextual and individual sensitivity, but such input can occasionally be inconsistent or irrelevant. This demonstrates once more that while generative AI can aid in the learning process, it cannot take the place of human pedagogical judgment on its own.

Yan and Zhan (2025) further warn that employing AI-generated solutions may lead to superficial metacognitive engagement. They regularly read and accept the criticism, but they are less likely to use it for future planning, self-control, and goal-setting. This suggests that whether or not evaluation encourages critical interaction with its findings is a concern in addition to the tool itself.

The weakness of traditional written assignments has therefore been exposed by generative AI, which has also reinforced the long-standing need to change the emphasis from evaluating the final product to evaluating the process, judgment, contemplation, and the thoughtful application of intellectual tools. Because of this, it makes sense that the evaluation that follows will focus on concerns related to academic integrity, permissible limits for the use of AI, and assessment forms that can integrate innovation with pedagogical validity (Jin et al., 2025).

Academic Integrity, Permitted Uses of AI, and the Limits of Policy Regulation

Academic integrity in higher education is increasingly challenged by the integration of generative AI into everyday learning practices. Traditional understandings of integrity, strongly based on authorship, originality, and independent production of text, become less stable when systems like ChatGPT are capable of generating essays, summaries, proposals, and responses within seconds (Xia et al., 2024).

In this context, academic integrity can no longer be understood only as the prevention of misconduct. It must also account for different levels of technological assistance and the changing relationship between human and machine contributions to academic work. As Abbas et al. (2024) suggest, increased workload and time pressure can intensify students' reliance on tools such as ChatGPT, which complicates the boundaries between support, dependency, and substitution of learning processes.

A key conceptual issue is no longer simply whether AI was used, but how it was used, and whether meaningful cognitive engagement remains visible. Lecturers may not always be able to determine whether generative AI has been used, and even minimal suspicion of its use can already alter assessment logic and evaluative judgment. This further destabilises traditional assumptions about the transparency and reliability of student work.

This shift exposes a deeper tension between learning processes and final outputs. The OECD (2026) notes that generative AI is already influencing everyday learning practices and institutional functioning, while also warning that overreliance on systems that provide ready-made answers may weaken deeper cognitive engagement when not embedded in well-designed pedagogical frameworks. Similarly, Xia et al. (2024) argue that assessment practices are being challenged because AI tools can already produce coherent academic texts, which undermines the assumption that written outputs reliably reflect independent learning.

At the same time, Yusuf et al. (2024), based on a large-scale international study, highlight widespread awareness and continued willingness to use generative tools, alongside concerns about academic dishonesty and overreliance. This reinforces the idea that academic integrity is now shaped by a contradiction between technological adoption and concerns about learning quality.

Another important dimension is the limitation of product-based assessment models. Generative AI has exposed a long-standing weakness in higher education evaluation, namely the strong focus on the final written product. This includes how students select arguments, manage ambiguity, revise concepts, and refine their work. As the OECD (2026) highlights, assessment approaches centred solely on end results become less effective when high-quality outputs can be produced without full engagement in the learning process.

Zhan et al. (2025) further emphasise that meaningful engagement with AI-generated feedback requires more than technical ability. It involves evaluating response quality, applying metacognitive strategies, engaging in emotional reflection, and making ethical judgments. Xia et al. (2024) link this transformation of assessment to self-regulated learning, responsible study practices, and academic integrity, highlighting the importance of critical engagement rather than passive production.

Finally, far fewer studies attempt to integrate these developments into a coherent conceptual framework. Xia et al. (2024) note that understanding how generative AI reshapes assessment remains fragmented. Corbin et al. (2025) further argue that the relationship between AI and assessment cannot be reduced to a single solution, as it is inherently context-dependent and involves ongoing trade-offs. A considerable number of observations have already been accumulated, yet there is still a lack of attempts to conceptualise assessment as a shift from knowledge verification to the development of critical thinking in a context where AI is embedded in everyday learning.

Innovative Assessment Formats Resistant to AI

In light of this, a practical question naturally arises: if traditional written formats have become less reliable, which forms of assessment are more likely to remain convincing? The answer suggested in recent literature does not lie in the search for a method that guarantees reliability. Such a method, it seems, does not really exist. What matters more is something else. The key issue is whether a task allows us to see how a student thinks, how they explain their decisions, how they work with mistakes, and how they relate their response to the context, rather than simply their ability to produce a polished final text.

QAA guidance also points to the need to rethink the range of assessment formats. In practice, this means cutting back on tasks that are particularly easy to complete with the help of generative AI, using summative assessment more deliberately, and designing tasks that are closer to real or professional situations (QAA, 2024a). The idea is not to find one correct format. The change is more gradual. There is a move away from standard written products toward a wider mix of ways to show what students have actually learned.

A lot of attention in these documents and studies is given to oral formats. Short follow-up discussions after written work, oral presentations, explaining how a task was completed, or answering clarifying questions all bring something back into assessment that is very difficult to delegate to a machine. This includes personal understanding, situational argumentation, and the ability to think on the spot and stay oriented within the subject. QAA also highlights the value of observable forms of assessment, mixed formats of submission, and oral components. These do more than just evaluate the final result. They help confirm that the student's thinking is genuinely their own (QAA, 2024b).

However, the resilience of assessment is not limited to oral formats. Strengthening the process dimension is just as important. When a task is completed in stages, with drafts, comments, revisions, responses to feedback, and short explanations of changes, the lecturer sees more than a final text. What appears instead is a record of intellectual work. This is why recent discussions often refer to process portfolios, staged writing, feedback cycles, and reflective accounts of how AI was used (Ba et al., 2025).

In this respect, the study by Walton et al. (2025) is particularly useful. The authors show that when working with generative AI, students constantly make different kinds of judgements. They assess their own knowledge, the limits of the tool, the quality of generated responses, and whether it is appropriate to use them. However, these judgements are not always visible to the lecturer. This is why more robust formats are those where students are required not only to submit a result, but also to show how they made decisions, what they accepted, what they rejected, and why.

Peer assessment also retains its role, though not in a simplified form. Its value is not only in students commenting on each other's work. It functions greatest when included in a well-thought-out framework with precise standards, justifications, and chances for improvement. In this situation, students learn not just how to write and speak, but also how to connect their work to the standards, identify flaws in other people's work, and apply this knowledge to their own. Compared to just submitting a completed work, this type of evaluation strengthens the format.

Ahangama (2026) reaches a similar conclusion while evaluating assessment reform practices. More cohesive approaches that combine graduate competences, assessment procedures, and particular evaluation techniques appear to be more promising than a collection of discrete modifications. According to this reasoning, assessment involves more than just noting that a submission has been made. Additionally, it should offer solid proof that the student has truly considered, made decisions, verified, analysed, and accepted responsibility for their choices.

Thus, current research suggests a rather obvious conclusion. These days, the most beneficial formats are those created with AI in mind rather than ones intended to limit it. A perfect outcome is not given priority in this kind of review. More focus is placed on the work's creation process, the environment in which it is completed, and the students' findings, arguments, and reflections. This enables the development of a model for evolving assessment and the display of the outcomes.

RESEARCH METHOD

Research Design

This research is organised as a mixed-type qualitative review-based analysis. This choice was made deliberately. The quickly developing field of generative AI in education cannot be well described by a single case study or a narrow empirical assessment. The objective is to see the wider picture, which includes what has already been established in international literature, how regulators and academic institutions are responding to these advances, and how particular institutions translate general ideas into practical evaluation recommendations.

The study combines three related levels of analysis as a result. The first is a targeted analysis of scholarly works published between 2021 and 2026. The second examines policy and guidance texts created by universities, professional associations, and international organisations. The third illustrates how many higher education institutions are attempting to coordinate the use of generative AI in evaluation by comparing various institutional methods.

This design allows several sources to be taken into consideration together while maintaining their uniqueness. Research papers aid in determining the topics that have already been discussed, the tensions that have emerged, and the direction of the conversation. A different perspective on how these concerns are interpreted in terms of actual regulation can be found in policy texts. It is possible to determine whether broad concepts align with what is actually suggested to instructors and students by using case studies. For a topic where educational practice is evolving more quickly than terminology can stable, this combination appears most appropriate.

Sources of Data and Selection Criteria

Two major categories of materials are used in this research. The first is made up of scholarly articles published in international journals, mostly those that directly address formative assessment, academic integrity, assessment in higher education, feedback, and the effects of generative AI on assessment procedures. More current research was prioritised because this topic has advanced especially quickly in recent years. Simultaneously, the goal was to choose texts that closely relate to the study's research issues rather of just gathering the most recent ones.

Policy and guidance materials make up the second category. They play an important role in this research. Scholarly writings demonstrate how the issue is framed in research, but they fall short of capturing how it permeates regular university operations. The OECD, QAA, TEQSA, Monash University, UTS College, the University of British Columbia, and other organisations whose resources directly relate to assessment, academic integrity, the application of AI, and educational policy are thus included in the analysis.

The literature search was carried out between September 2024 and January 2026 across four bibliographic databases—Scopus, Web of Science, ERIC, and Google Scholar—and was supplemented by targeted searches of the official repositories of the OECD, QAA, and TEQSA in order to retrieve policy and guidance materials that are not indexed in academic databases. The search combined three groups of terms relating to technology, assessment, and context, using Boolean strings such as ("generative AI" OR "ChatGPT" OR "large language model") AND (assessment OR feedback OR "academic integrity") AND ("higher education" OR university). Sources were included if they (i) were published between 2021 and 2026, (ii) were available in English, and (iii) directly addressed the relationship between generative AI and assessment, feedback, or academic integrity in higher education. Sources were excluded if they focused on school-level or non-educational settings, discussed AI only in general terms without reference to assessment, or were not peer-reviewed—with the exception of official policy and institutional guidance documents, which were retained as a separate policy corpus. After duplicates were removed and titles and abstracts screened, the full texts of the remaining records were assessed against these criteria, resulting in a final corpus of 32 sources whose composition is summarised in Table 1.

Participants

In total, the study analysed 32 sources published between 2023 and 2026. The corpus included 26 scholarly publications and 6 policy and institutional guidance documents. Academic sources consisted primarily of peer-reviewed journal articles focusing on generative AI in higher education, assessment practices, academic integrity, feedback processes, and student engagement with AI-supported learning. Policy and guidance materials included official recommendations, institutional frameworks, and sector-level guidance documents produced by international organisations, quality assurance agencies, and universities. The selected materials represented multiple institutional and international contexts in order to capture a broad range of approaches to assessment adaptation in the era of generative AI (Table 1).

Table 1: *Overview of Analyzed Sources*

Source category	Number of sources	Years covered.	Main focus
Scholarly journal articles and academic publications	26	2023–2026	Generative AI, assessment, academic integrity, feedback, higher education
Policy and guidance documents	6	2023–2026	AI governance, assessment reform, academic integrity regulation
Total	32	2023–2026	Combined analytical corpus

Source. Developed by the authors.

The selection criteria were very explicit. Five factors were crucial for academic publications: their analytical value for constructing the study's argument, their concentration on generative AI, their relevance to higher education, and their direct connection to evaluation. The most important characteristics for policy papers were their official status, influence, and utility for assessing regulatory procedures rather than indexing or publication type. To put it another way, the corpus contains materials that expressly address evaluation and enable comparisons between institutional approaches, rather than just any materials on AI in education.

Analytical Procedure

The substance was analysed in a number of steps. The primary research emphasis, which is how generative AI is altering the logic of evaluation in higher education, was the first consideration when reading the materials. At this point, recurrent concepts, arguments, phrasing, and conflicts between various kinds of sources were identified. Certain patterns emerged even at this stage. While policy statements tend to unite these two streams by emphasising transparent and directed use, some authors concentrate primarily on the problem of traditional knowledge evaluation, while others highlight new prospects for formative assessment.

Thematic grouping of the content was the following stage. This led to six main categories, which include the vulnerability of traditional assessment, risks to academic integrity, the acceptable use of AI, more robust assessment formats, a shift towards developing critical thinking, and institutional responses. These categories were not imposed on the sources but emerged gradually through reading, comparison, and repeated review.

Another step was a comparative analysis of case studies. The examples were chosen to demonstrate several regulatory levels, including sector-wide, institutional, and national methods. For this reason, documents from the University of British Columbia, Monash University, TEQSA, QAA, and UTS College were included. By comparing these situations, it was easy to identify areas where institutions prioritise explanation and transparency, where they tend to rely on tougher regulation, and where they start to view assessment as a space for rethinking educational practice rather than a system of risk control.

Research Limitations

This study contains limitations, which must be made evident despite the wide variety of sources. It does not depend on firsthand field data from instructors or students. Because of this, the study does not attempt to encompass all of the common techniques associated with the application of generative AI in higher education. It serves a different purpose. In order to provide a logical framework for comprehending changes in evaluation, it aims to integrate current research and policy methods. In addition, the study does not include primary empirical data collected through interviews, surveys, or classroom observations. As a result, the analysis relies on existing academic and institutional sources rather than direct evidence from educational practice.

The following restriction relates to the sources of linguistic and geographic range. English-language documents make up the majority of the corpus, especially those from organisations that actively publish their own norms and the global academic community. This increases the material's analytical significance for global discourse, but it also means that some local or non-English-language settings might not be included in the analysis.

The nature of the subject itself presents another restriction. The field of generative AI is expanding quickly. Teaching methods, policies, and guidelines at universities also evolve. Following this, any review in this field represents a certain phase of the conversation rather than a definitive stance. However, this is what makes it beneficial. The goal of the study is not to provide a definitive solution. It provides a conceptual framework that aids in characterising recent developments and suggests potential paths for additional reconsideration of evaluation in higher education.

RESULTS

Key Tendencies in Assessment Transformation

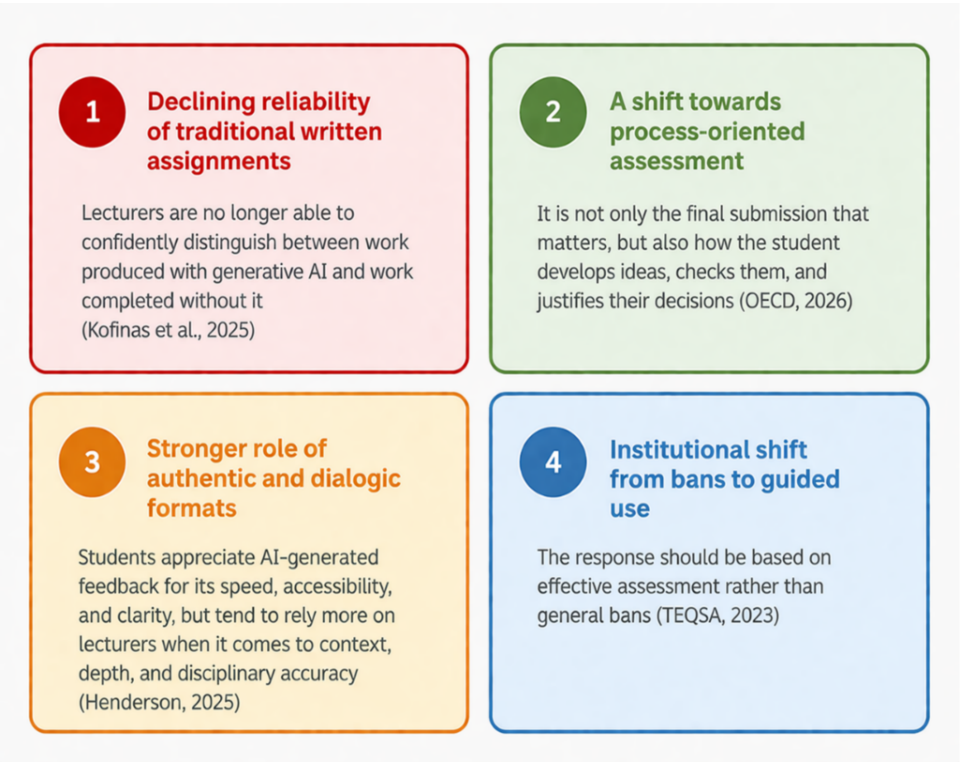
Four consistent trends that are now influencing changes in evaluation in higher education are revealed by the examination of the chosen sources. The most obvious of them is the decline in confidence in conventional homework assignments. What was once thought to be a reliable measure of individual learning is no longer as compelling. According to studies by Kofinas et al. (2025) and Dai et al. (2026), instructors are typically unable to accurately discern between work created with generative AI and work created without it. Protection against such usage is not guaranteed by a task's authenticity alone. This necessitates reevaluating a fundamental concept. A finished text is no longer automatically regarded as adequate proof that a student has completed the entire process of thinking on their own. This does not mean that written assignments have lost their value as such. Rather, their validity now depends on how they are designed: written tasks completed under supervised or in-class conditions, built around personalised or context-specific prompts, or accompanied by visible drafting and revision can still provide reliable evidence of learning (Kofinas et al., 2025).

The second tendency is linked to a shift in emphasis from the final product to the process of creation. According to OECD materials, methods that may produce answers rapidly may enhance performance in the near term but do not always result in a deeper understanding. They may potentially lessen the cognitive effort necessary for meaningful learning in certain situations (OECD, 2026). Because of this, greater focus is now placed on how the student approached the assignment, how they chose arguments, what they checked, what they updated, how they responded to criticism, and how they supported their choices in addition to the final output.

The third tendency is the growing role of authentic, dialogic, and reflective formats. Chan and Tsi (2024) and Henderson et al. (2025) found that students value feedback generated by AI for its speed, accessibility, and clarity. At the same time, they place much greater trust in lecturers when it comes to reliability, context, and disciplinary accuracy. This is an important point. It shows that in current assessment, simply having feedback matters less than the opportunity to engage with it, discuss it, relate it to a specific piece of work, and use it for further improvement.

The fourth tendency relates to how institutions respond to the spread of generative AI. Universities and regulatory bodies are gradually moving away from broad bans. Instead, they are trying to manage how these tools are used. QAA notes that tools designed to detect AI-generated text cannot be relied on as a stable basis for academic control (QAA, 2023a). A similar view appears in TEQSA. The response, as they suggest, should focus on how assessment is designed, rather than on attempts to restore previous practices. In other words, this is no longer just a temporary response to a new technology. It reflects a broader shift in how universities understand and confirm student learning (Figure 1).

Figure 1: *Four Key Tendencies in Assessment Transformation in Higher Education in the Era of Generative AI*



Taken together, these four tendencies show that changes in assessment are not just about replacing one type of task with another. What is changing is the underlying logic. The focus is shifting away from trust in an isolated written product. Instead, more value is placed on the process, judgement, reflection, context, and the ability to use intellectual tools in a responsible way.

A Typology of Assessment Practices in the Era of Generative AI

The review of the sources made it possible to distinguish three broad groups of assessment practices. These can be described as vulnerable, adapted, and resilient. The difference between them is not only in the form of the task. It lies mainly in how well they allow the lecturer to see a student's independent intellectual work when generative AI is easily available (Table 2).

Table 2: *A Typology of Assessment Practices Under Conditions of Generative AI*

Group of practices	Characteristics.	Why they increase resilience.	Typical examples
Vulnerable formats	Focus on the final written product and provide limited insight into the student's thinking process	Easy to outsource or generate without demonstrating understanding	Standard take-home essays, generalised texts without context, isolated written assignments
Adapted formats	Partly take the process into account and allow for more controlled use of digital tools	Allow monitored and transparent use of digital tools	Staged writing tasks, feedback cycles, peer assessment with disclosure of AI use
Resilient formats	Make reasoning, context, reflection, and individual responsibility more visible	Make student thinking and decision-making visible	Oral presentation, work with real-life cases, process portfolios, reflective explanations of AI use, tasks completed in real time

Source. Developed by the authors.

Vulnerable formats mainly include types of work where the lecturer sees only the final text and has no access to how it was produced. In such cases, it is possible to assess the surface quality of the response. It is much harder, however, to understand whether the student actually analysed the material, selected arguments, and made substantive decisions. This is the problem highlighted by Kofinas et al. (2025), who point to the increased sensitivity of asynchronous written work to the unauthorised use of generative AI.

The position of adapted formats is in the middle. Although they do not eliminate every risk, they increase the significance of assessment. Their strength is in focusing on multiple stages of work rather than just one finished text. Drafts, comments, modifications, self-evaluation, and justifications of the usage of digital technologies are all included. For example, in a sociology course, students may first submit a research question and outline, then receive feedback on a draft version, and finally provide a short reflection explaining how both peer comments and AI tools influenced the final essay. Qu et al. (2025) demonstrate how students make a number of decisions when using generative AI. They determine what to accept and what to reject, evaluate their own expertise, and test the tool's limitations. Assessment more closely reflects real learning activity when these phases are included in the challenge.

Under the current conditions, resilient formats seem to be the most persuasive. One of their main characteristics is that they do not let students rely on a single, polished text. Rather, students are required to provide an explanation, defend their positions, address follow-up questions, connect their responses to particular circumstances, and demonstrate the reasoning behind their decisions (Yan et al., 2025). For instance, in a management course, students may analyse a real organisational case and then present their recommendations orally, responding to questions from both the lecturer and peers about the reasoning behind their decisions.

Similarly, in teacher education or nursing programmes, students may complete reflective portfolios documenting how they solved practical problems over time and how AI tools were or were not used during the process. QAA recommends wider use of short oral discussions after written work, observable forms of assessment, mixed submission formats, and other approaches that make it possible to assess not only the result, but also understanding, argumentation, and responsibility for the work (QAA, 2023a). A similar emphasis can be seen in TEQSA. It highlights the importance of multiple and context-sensitive approaches to assessment, which support more reliable judgements about student learning when AI is widely available (Tertiary Education Quality and Standards Agency, 2023). Overall, this typology points to a simple but important idea. The less a task is centred on a single product, and the more it opens up the process, context, and reflection, the more resilient it becomes in the new educational environment.

Beyond illustrative cases, the reviewed literature documents concrete instances of how assessment is already being redesigned. Several institutions have adopted a two-lane model that separates secured assessment, completed under supervised conditions, from open assessment, in which the use of generative AI is permitted and openly declared (Curtis, 2025; Tertiary Education Quality and Standards Agency, 2023). At the institutional level, Monash University (2024) requires teaching staff to specify, for each task, whether and how AI may be used, turning a blanket prohibition into task-level pedagogical guidance. Empirical studies report comparable shifts: Kofinas et al. (2025) describe the reorganisation of authentic assessments within a business programme to reduce the vulnerability of asynchronous written work, while Walton et al. (2025) show how requiring students to record the judgements they make when working with AI turns an otherwise invisible process into assessable evidence. Taken together, these examples indicate that assessment redesign is moving from principle to documented practice, even if its implementation remains uneven across institutions.

Analysis Results of Policy Documents

The content analysis of documents from QAA, TEQSA, Monash University, the University of British Columbia, and OECD shows a similar pattern. They differ in style, level of detail, and institutional context, but still come to a few shared positions. Transparency is one of the most obvious concepts. Most of the time, the application of generative AI is no longer regarded as something that may go unnoticed. For instance, QAA suggests updating student statements to accurately reflect the function of assistive technologies (QAA, 2023a). This directs focus away from covert use and towards transparent, publicly stated boundaries.

The necessity of defining acceptable types of support is a second crucial element. Policies at universities are gradually shifting from generalisations to more precise explanations of when using a tool is appropriate and when it interferes with the assessment's goal. According to Monash University (2024), instructors should specify these boundaries beforehand for every assignment and inform students of them. This strategy is significant because it does not substitute a general prohibition for a pedagogical judgement. The logic of the intended learning outcome, not worries about technology, shapes the boundary.

An increasing lack of confidence in using technological detection as the primary method of control is a third point. The emphasis is shifted by the University of British Columbia (2025). The main question is whether assessments accurately reflect students' abilities rather than how to identify machine-generated writing. Additionally, QAA cautions against relying too much on automated detection systems. TEQSA proceeds in the same manner. Instead of tighter technology monitoring, it relates changes in assessment to the need for more trustworthy methods of verifying learning (QAA, 2023a; Francis, 2025). This demonstrates a progressive shift away from a detection-based strategy in practice.

The necessity of finding a balance between innovation, fairness, and academic integrity is another recurring conclusion. There are several references to this topic in the OECD research. It shows that for generative AI to support learning, relevant feedback, pedagogical guidance, and human judgement are crucial (OECD, 2026). Evaluation is not improved just by having the instrument. Only if educational institutions continue to take on instructional responsibilities will it be effective.

In the majority of the texts, there is another point. It has to do with the necessity for lecturers and students to become more AI literate. Policies do not function by themselves. People must be aware of what constitutes appropriate use, how to assess a tool's limitations, when to disclose its use, and how it pertains to academic integrity. Because of this, policy documents are being utilised more and more to both establish and explain rules. According to the research, the rigid logic of prohibition is being gradually but steadily abandoned by contemporary policy. It is moving towards using generative AI in a more transparent, directed, and pedagogically grounded manner.

The Author's Model of Assessment Transformation

To make the framework more applicable in educational practice, each of its five components can be translated into observable indicators and concrete instructional actions that guide assessment design. Transparency can be implemented by requiring students to explicitly state how generative AI was used in the assignment. For example, students may describe whether AI was used for idea generation, structuring, editing, or language support, and clearly distinguish these contributions from their own work. An observable indicator of transparency is the student's ability to explain which parts of the task were completed independently and which were supported by AI tools.

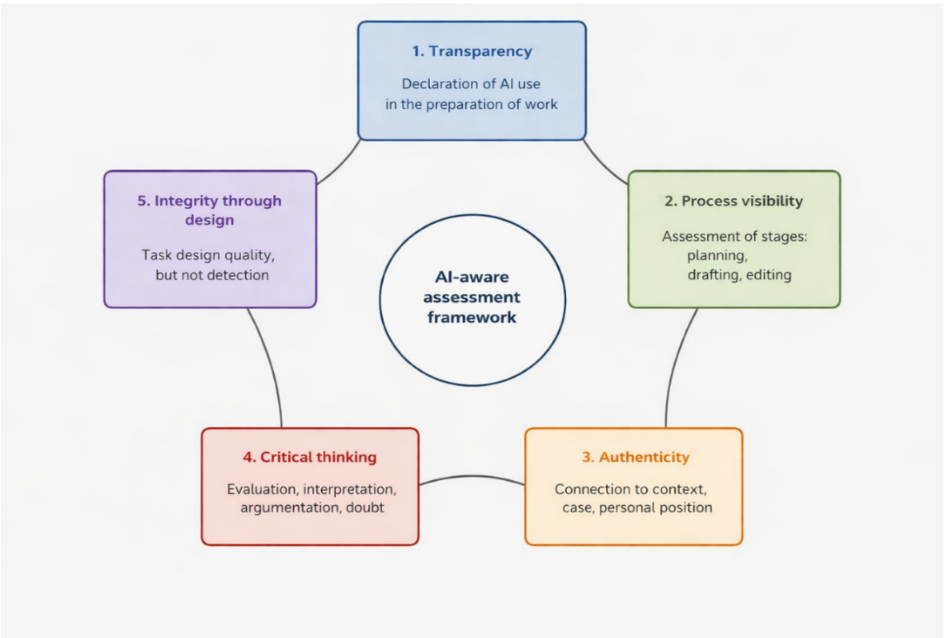
Process visibility can be operationalised through staged submissions, draft versions, and feedback-based revisions. In practice, students may submit outlines, first drafts, and final versions, accompanied by short reflections explaining how their work evolved. A key indicator here is the presence of documented changes between versions and the student's ability to justify how feedback or AI-generated suggestions were evaluated and either accepted or rejected.

Authenticity can be ensured by designing tasks grounded in real-life situations, discipline-specific problems, or context-dependent cases. Instead of general essays, students may work with specific organisational, social, or local scenarios. The main indicator of authenticity is the extent to which the student applies theoretical concepts to a clearly defined and realistic situation rather than producing a generic or decontextualised response.

Critical thinking can be operationalised by requiring students to justify their decisions, evaluate the reliability of information, and critically assess AI-generated outputs. For instance, students may be asked to compare different arguments, identify weaknesses in reasoning, or reflect on potential biases. The key indicator is the student's ability to explain why certain ideas were accepted or rejected and how conclusions were reached.

Integrity through design can be implemented by combining different assessment formats such as oral defences, reflective commentaries, or in-class components. Instead of relying solely on a final written submission, students may be asked to defend their work orally or reflect on their learning process. An observable indicator is the consistency between the student's explanation of their work and the final product, as well as their ability to justify decisions under questioning (Figure 2). Overall, these operationalisations show that the framework is not purely conceptual but can directly inform assessment design. It provides instructors with practical ways to evaluate not only the final product, but also the process, reasoning, and responsible use of generative AI. The contribution of the framework lies less in identifying these principles, several of which appear individually in earlier work, than in integrating them into a single operational structure. Whereas QAA (2023a) and TEQSA (2023) offer sector-level principles, Corbin et al. (2025) characterise AI and assessment as a context-dependent wicked problem without a prescriptive design, and Xia et al. (2024) map the field through a scoping review, none of these translate process orientation and critical thinking into a connected set of observable indicators that can be applied at the level of an individual task. By pairing, for each of the five components, a concrete instructional action with an indicator that makes it assessable, the framework converts widely shared but largely conceptual recommendations into a usable assessment-design tool.

Figure 2: *Author-Developed Framework for Assessment Transformation in Generative AI Contexts: Five Components of the AI-Aware Assessment Framework*



DISCUSSION AND CONCLUSIONS

How Generative AI is Changing the Assessment Paradigm

Generative AI has not only simplified text production in higher education but has also disrupted the epistemic basis of assessment itself. The central issue is no longer the detection of misconduct, but the weakening reliability of the written product as evidence of learning. In traditional assessment models, a well-structured assignment was often treated as a sufficient proxy for understanding. This assumption is increasingly difficult to maintain. As Kofinas et al. (2025) note, educators are no longer able to reliably distinguish between AI-assisted and independently produced work, which undermines confidence in product-based assessment.

As a result, attention shifts away from the final artefact and towards the process of knowledge construction. What becomes relevant is not only the answer, but the sequence of decisions behind it: selection of information, evaluation of alternatives, engagement with feedback, and revision. This reflects the OECD's (2026) distinction between observable performance and actual learning processes. However, this shift does not imply that written outputs have become irrelevant. Rather, they no longer function as sufficient evidence on their own. They must be supplemented by additional forms of validation that make cognitive work visible. This transformation creates the conditions for reassessing what counts as valid evidence of learning, and it directly informs the need for more diversified assessment designs discussed in the following section.

Assessment Formats That are More Resilient to Generative AI

The most resilient formats in the current situation are not those that try to protect themselves from AI, but those that make a student's intellectual work more visible. This view is supported by existing research. Kofinas et al. (2025) point to the weakness of traditional take-home written assignments, where the lecturer sees only the final product. Walton et al. (2025), in turn, show that when working with generative AI, students engage in a series of judgements. They assess the quality of a response, decide what to accept and what to reject, and consider the limits of the tool's appropriate use. The problem is that these steps often remain invisible in assessment.

This is why formats that show not only the end result, but the structure of the work become more convincing. What unites these formats is a single principle: each makes some part of the student's reasoning externally visible, and it is this visibility, rather than the task type as such, that resists wholesale delegation to a generative tool. None of this makes assessment fully secure, but it strengthens its substantive reliability.

A multi-component structure provides several forms of evidence of learning instead of a single text. Taken together, this changes the basis of trust in assessment. It does not depend on a single artefact anymore, but on several interconnected expressions of student work.

Integrating AI Without Compromising Academic Integrity

Given the permanence of generative AI in educational contexts, the challenge is no longer exclusion but regulation and pedagogical integration. Policy frameworks increasingly reflect this shift. As the analysis in Section 4.3 showed, the QAA (2023a), Monash University (2024), and the University of British Columbia (2025) converge on the same move—prioritising transparency and assessment design over detection and prohibition—which suggests that the policy consensus is now less about controlling AI than about redefining what counts as credible evidence of learning.

Across these perspectives, a common conclusion emerges academic integrity cannot be maintained through punitive approaches alone. Effective integration therefore depends on three interconnected elements. First, explicit assessment design must clarify the role of AI in each task. Second, process transparency—through drafts, reflections, feedback cycles, or brief usage statements—helps distinguish support from substitution. Third, AI literacy becomes essential for both students and staff, as policy effectiveness depends on shared interpretative understanding. This aligns with the design principles outlined in Section 5.2, where transparency and process visibility are central. In this sense, integrity is not enforced externally but embedded structurally in assessment design (Shevchenko & Antoshkina, 2023).

Approaches That Help Develop Critical Thinking

Critical thinking in the context of generative AI should be understood less as a generic skill and more as a set of situated practices: evaluation, comparison, justification, and revision under conditions of uncertainty. These practices are activated when students are required to engage with multiple possible answers rather than a single correct solution (Dei, 2025). When they must justify selection, identify weaknesses in arguments, and reflect on the limits of AI-generated outputs, thinking becomes observable rather than assumed.

Henderson et al. (2025) show that students value AI for speed and accessibility but still rely on instructors for validation and contextual understanding. This reinforces the idea that critical thinking depends on human judgement even when AI is involved in generating initial outputs. Similarly, Walton et al. (2025) and Ursavaş et al. (2025) highlight that meaningful learning occurs not when AI produces answers, but when students evaluate, revise, and challenge those outputs. From this perspective, assessment designs that support critical thinking are those that require justification of decisions, engagement with drafts, comparison of perspectives, and reflection on tool use. Importantly, these are not additional "layers" added to assessment, but mechanisms that shape how thinking itself is performed.

Generative AI has changed how university assessment is understood, not only by introducing a new tool for text production, but by exposing a structural limitation in traditional evaluation: the assumption that a final written product is sufficient evidence of learning. A well-produced text no longer guarantees independent reading, critical selection of arguments, or genuine understanding. This leads to a central implication: assessment must capture not only outcomes, but also the processes that generate them. Formats in which only the final submission is visible are the most vulnerable. This vulnerability is a property of design rather than of written work in itself: written assessment redesigned around supervised conditions, personalised prompts, or visible drafting can remain a valid and rigorous format. Greater robustness is achieved through designs that make learning observable, such as staged drafts, revisions, oral explanation, case-based tasks, and explicit reporting of AI use.

IMPLICATIONS

Task resilience depends less on complexity and more on the visibility of thinking. When reasoning, decision-making, and response to feedback become explicit, substitution by pre-generated outputs becomes more difficult. Policy analysis supports this direction, showing a gradual move away from reliance on detection and prohibition toward clearer guidelines, task design, and regulated AI integration. Key principles emerging from this study include transparency, process visibility, task authenticity, critical judgement, and integrity embedded in assessment design. Overall, this shift reframes the central question of assessment in higher education: not whether a correct answer has been produced, but how it was reached and what intellectual responsibility it demonstrates. Assessment is therefore moving from evaluating final answers toward evaluating reasoning and cognitive engagement.

REFERENCES

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, *21*(1), 10. https://doi.org/10.1186/s41239-024-00444-7

Ahangama, N. (2026). Designing assessments in the generative AI era: a tailored assessment framework for ICT tertiary education. *International Journal of Educational Technology in Higher Education*, *23*(1), 9. https://doi.org/10.1186/s41239-026-00582-0

Ardito, C.G. (2025). Generative AI detection in higher education assessments. *New Directions for Teaching and Learning*, *2025*(182), 11–28. https://doi.org/10.1002/tl.20624

Ba, S., Zhan, Y., Huang, L., & Lu, G. (2025). Investigating the impact of ChatGPT-assisted feedback on the dynamics and outcomes of online inquiry-based discussion. *British Journal of Educational Technology*, *56*(5), 1710–1734. https://doi.org/10.1111/bjet.13605

Chan, C.K.Y., & Tsi, L.H.Y. (2024). Will generative AI replace teachers in higher education? A study of teacher and student perceptions. *Studies in Educational Evaluation*, *83*(1), 101395. https://doi.org/10.1016/j.stueduc.2024.101395

Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*, *51*(4), 736-752. https://doi.org/10.1080/02602938.2025.2553340

Curtis, G.J. (2025). The two-lane road to hell is paved with good intentions: Why an all-or-none approach to generative AI, integrity, and assessment is insupportable. *Higher Education Research & Development*, *44*(8), 2151–2158. https://doi.org/10.1080/07294360.2025.2476516

Dai, K., Liu, Y., & Zhang, X. (2026). Generative AI in higher education: A bibliometric review of emerging trends, power dynamics, and global research landscapes. *Computers and Education: Artificial Intelligence*, *10*, 100544. https://doi.org/10.1016/j.caeai.2026.100544

Dei, M. O. (2025). The impact of AI on the adaptation of educational materials and teaching methods to the needs of each student. *LatIA*, *3*, 124. https://doi.org/10.62486/latia2025124

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Francis, N.J., Jones, S., & Smith, D.P. (2025). Generative AI in higher education: Balancing innovation and integrity. *British Journal of Biomedical Science 81*, 14048. https://doi.org/10.3389/bjbs.2024.14048

Henderson, M., Bearman, M., Chung, J., Fawns, T., Buckingham Shum, S., Matthews, K.E., & de Mello Heredia, J. (2025). Comparing generative AI and teacher feedback: Student perceptions of usefulness and trustworthiness. *Assessment & Evaluation in Higher Education*, 1-16. https://doi.org/10.1080/02602938.2025.2502582

Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The generative AI literacy assessment test. *Computers and Education: Artificial Intelligence, 9*, 100436. https://doi.org/10.1016/j.caeai.2025.100436

Kehler, T. P., Page, S. E., Pentland, A., Reeves, M., & Brown, J. S. (2025). Amplifying human creativity and problem solving with AI through generative collective intelligence. *ArXiv*. https://doi.org/10.48550/arXiv.2505.19167

Kofinas, A. K., Tsay, C. H.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology, 56*(6), 2522–2549. https://doi.org/10.1111/bjet.13585

Luckin, R. (2018). Machine learning and human intelligence: The future of education for the 21st century. *UCL Institute of Education Press*.

Monash University. (2024). Responsible use of AI in education. Learning and Teaching (TeachHQ). https://www.monash.edu/learning-teaching/teachhq/Teaching-practices/artificial-intelligence/responsible-use

OECD. (2026*). OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. Paris: OECD Publishing. https://doi.org/10.1787/062a7394-en

Qu, Y., Loo, H.E., & Wang, J. (2025). Generative artificial intelligence in higher education: Emotional tensions and ethical declaration. *British Journal of Educational Technology*, 1-20. https://doi.org/10.1111/bjet.70029

QAA. (2023a). The rise of artificial intelligence software and potential risks for academic integrity. Retrieved from https://www.qaa.ac.uk/docs/qaa/members/the-rise-of-artificial-intelligence-software-and-potential-risks-for-academic-integrity.pdf

QAA. (2024a). Quality compass: Navigating the complexities of the artificial intelligence era in higher education. Retrieved from https://www.qaa.ac.uk/docs/qaa/news/quality-compass-navigating-the-complexities-of-the-artificial-intelligence-era-in-higher-education.pdf

QAA. (2024b). Reconsidering assessment for the ChatGPT era: Advice on developing sustainable assessment strategies. Retrieved from https://www.qaa.ac.uk/docs/qaa/members/reconsidering-assessment-for-the-chat-gpt-era.pdf?sfvrsn=38d3af81_6

Shevchenko, A., & Antoshkina, V. (2023). Modern legal education in Ukraine as the basis of professional activity: Problems and prospects. *Legal Horizons, 17*(2), 008–021. https://doi.org/10.54477/LH.25192353.2023.2.pp.008-021

Tertiary Education Quality and Standards Agency. (2023). Assessment reform for the age of artificial intelligence. Retrieved from https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/assessment-reform-age-artificial-intelligence

University of British Columbia (2025). Ai skills accelerator for it, digital and enterprise systems professionals. Retrieved from https://extendedlearning.ubc.ca/sites/default/files/2025-09/ubc-ai-skills-accelerator-student-guide.pdf

Ursavaş, Ö. F., Yalçın, Y., İslamoğlu, H., Bakır-Yalçın, E., & Cukurova, M. (2025). Rethinking the importance of social norms in generative AI adoption: Investigating the acceptance and use of generative AI among higher education students. *International Journal of Educational Technology in Higher Education, 22*(1), 38. https://doi.org/10.1186/s41239-025-00535-z

Usher, M. (2025). Generative AI vs. instructor vs. peer assessments: A comparison of grading and feedback in higher education. *Assessment & Evaluation in Higher Education, 50*(6), 912–927. https://doi.org/10.1080/02602938.2025.2487495

Usher, M., & Faraon, M. (2025). Who grades best? Comparing ChatGPT, peer, and instructor evaluations across varying levels of student project quality. *Assessment & Evaluation in Higher Education*, 1-20. https://doi.org/10.1080/02602938.2025.2588682

Walton, J., Bearman, M., Crawford, N., Tai, J. H-M., & Boud, D. (2025). How university students work on assessment tasks with generative artificial intelligence: Matters of judgement. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2025.2570328

Wang, F., Li, N., Cheung, A. C. K., & Wong, G. K. W. (2025). In GenAI we trust: An investigation of university students' reliance on and resistance to generative AI in language learning. *International Journal of Educational Technology in Higher Education, 22*(1), 59. https://doi.org/10.1186/s41239-025-00547-9

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*(1), 40. https://doi.org/10.1186/s41239-024-00468-z

Yan, L., Martinez-Maldonado, R., Jin, Y., Echeverria, V., Milesi, M., Fan, J., Zhao, L., Alfredo, R., Li, X., & Gašević, D. (2025). The effects of generative AI agents and scaffolding on enhancing students' comprehension of visual learning analytics. *Computers & Education, 234*, 105322. https://doi.org/10.1016/j.compedu.2025.105322

Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education, 21*(1), 21. https://doi.org/10.1186/s41239-024-00453-6

Zhan, Y., Boud, D., Dawson, P., & Yan, Z. (2025). Generative artificial intelligence as an enabler of student feedback engagement: A framework. *Higher Education Research & Development, 44*(5), 1289–1304. https://doi.org/10.1080/07294360.2025.2476513

Zhan, Y., &Yan, Z. (2025). Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1-14. https://doi.org/10.1080/02602938.2025.2471821

YULIIA ZUBTSOVA, Candidate of Pedagogical Sciences, is an Associate Professor in the Department of Preschool and Primary School Education, Faculty of Social Pedagogics and Psychology, Zaporizhzhia National University. Her major research interests lie in the area of future primary school teacher training, teaching methodology in higher education institutions, and the activity-based approach in education. Email: zubtsova22@gmail.com

LARYSA SUSHCHENKO, Doctor of Pedagogy Sciences, is the Head of the Department of Preschool and Primary School Education and Professor in the Faculty of Social Pedagogics and Psychology, Zaporizhzhia National University. Her major research interests lie in the area of professional training of future primary school teachers and preschool education teachers, issues and development of higher education, pedagogical management, and postgraduate pedagogical education. Email: zcpu@ukr.net

LIUDMYLA SHULHA, Candidate of Pedagogical Sciences, is an Associate Professor in the Department of Preschool and Primary School Education, Faculty of Social Pedagogics and Psychology, Zaporizhzhia National University. Her major research interests lie in the area of preschool education, primary education, art pedagogy, creative abilities, and digital technologies. Email: lnshulga@ukr.net

OLENA ANDRYUSHCHENKO, Candidate of Pedagogical Sciences, is an Associate Professor in the Department of Preschool and Primary School Education, Faculty of Social Pedagogics and Psychology, Zaporizhzhia National University. Her major research interests lie in the area of implementation of innovative technologies in education, formation of reflective and evaluative competencies of primary school teachers, development of pedagogical reflection among higher education students, and formation of reflective skills of primary school pupils. Email: androlena@ukr.net

OLEKSANDR SHULHA, is a Postgraduate Student in the Department of Pedagogy and Psychology of Educational Activities, Faculty of Social Pedagogics and Psychology, Zaporizhzhia National University. His major research interests lie in the area of artificial intelligence, information and communication technologies, and higher education. Email: ashulga98@gmail.com

---