Medición y Evaluación del Aprendizaje Inclusivo. 2025; 2(2):19
doi: 10.63688/aprendizaje.v2.i2.19

REVISIÓN SISTEMÁTICA

# Effectiveness of generative artificial intelligence in formative assessment and feedback of learning in university students

## Efectividad de la inteligencia artificial generativa en la evaluación formativa y retroalimentación del aprendizaje en estudiantes universitarios

**Luis David Bastidas González**[1]

[1] Instituto Sapiens de Investigación Científica Multidisciplinar, Milagro-Ecuador

Submitted: 13-05-2025 | Revised: 09-06-2025 | Accepted: 29-08-2025 | Published: 04-09-2025

**Corresponding Author:** Luis David Bastidas González

**ABSTRACT**

Introduction: Efectividad de la inteligencia artificial generativa en la evaluación formativa y retroalimentación del aprendizaje en estudiantes universitarios is an emerging higher-education issue related to measurement quality, decision-making and learning experience. Objective: To systematically synthesize available evidence and identify patterns of effectiveness, methodological quality and implementation conditions. Method: A PRISMA-oriented systematic review was conducted on a consolidated academic corpus of 135 records; 24 studies met relevance and documentary sufficiency criteria. Results: Evidence was methodologically heterogeneous but converged on the need to align technology, constructs, feedback and educational decisions. Conclusions: Findings support evidence-informed adoption, institutional monitoring and continuous assessment of validity, equity and utility.

**Keywords:** higher education; systematic review; educational assessment; evidence; PRISMA; educational technology.

**RESUMEN**

Introducción: Efectividad de la inteligencia artificial generativa en la evaluación formativa y retroalimentación del aprendizaje en estudiantes universitarios representa un problema emergente para la educación superior por su relación con la calidad de la medición, la toma de decisiones y la experiencia de aprendizaje. Objetivo: sintetizar sistemáticamente la evidencia disponible y establecer patrones de efectividad, calidad metodológica y condiciones de implementación. Método: se aplicó una revisión sistemática con lógica PRISMA sobre un corpus académico consolidado de 135 registros, del cual se incluyeron 24 estudios que cumplieron criterios de pertinencia y suficiencia documental. Resultados: la evidencia muestra heterogeneidad de diseños y contextos, junto con convergencias en torno a la necesidad de alinear tecnología, constructo, retroalimentación y decisiones educativas. Conclusiones: los resultados respaldan una adopción basada en evidencia, seguimiento institucional y evaluación continua de validez, equidad y utilidad.

**Palabras clave:** educación superior; revisión sistemática; evaluación educativa; evidencia; PRISMA; tecnología educativa.

© 2026, The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0).

## INTRODUCTION

In higher education, research on the effectiveness of generative artificial intelligence in formative assessment and feedback on learning among university students has become increasingly relevant because assessment decisions no longer rely solely on static instruments or face-to-face exchanges. Digitalization has expanded the volume of evidence available on performance, interaction, and learning, but has also heightened the need to distinguish between technically attractive innovations and practices capable of supporting valid educational inferences. This review addresses this challenge from an integrative perspective, focusing on both observed effects and the methodological conditions that allow interpretation of these effects without confusing technological availability with assessment quality.

The conceptual core of this manuscript is organized around feedback mechanisms. This choice avoids treating the literature as a simple collection of results and makes it possible to examine how each study defines the phenomenon, what population it observes, what instruments it employs, and what type of conclusion it considers legitimate. The heterogeneity of university contexts requires attention to the relationship among design, implementation, and evidence, because the same strategy can produce benefits under certain conditions and modest results when instructional support, discipline, digital experience, or the feedback form changes.

This systematic review is relevant because the field combines experimental studies, instrumental validations, observational analyses, technological developments, and previous syntheses. This diversity makes it difficult to draw solid conclusions through fragmentary reading. Instead of assuming that all works respond to the same question, this article reconstructs patterns of convergence and divergence, identifies which results are repeated with greater consistency, and separates claims directly supported by studies from those that still require additional validation. The purpose is to offer a useful reading for research, teaching, and university management.

### Fundamentals of generative feedback

The research (1), published in 2025, provided evidence through exploratory mixed methods. The study used a survey (n = 40), scripts (n = 6), and interviews (n = 2), which makes it possible to situate the scope of its inferences and assess its proximity to the university population of interest. Substantively, the findings showed that collaborative and reflective use favored phases of cognitive presence, whereas passive AI-directed use resulted in less critical elaboration. This antecedent is relevant to the perspective of feedback mechanisms because it demonstrates that the interpretation of the effect depends on the correspondence among construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

Study No. (2), published in 2026, provided evidence through an empirical comparison of models. The study used essays from gender, ELL, and special education subgroups, which makes it possible to situate the scope of its inferences and assess its proximity to the target university population. Substantively, the findings showed that fine-tuned gPT-4o achieved the greatest accuracy and overall equity among the compared models. This antecedent is relevant to the perspective of feedback mechanisms because it demonstrates that the interpretation of the effect depends on the correspondence among construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The evidence synthesis (3), published in 2025, contributed evidence from a systematic literature review. It included 158 studies (2021–2024), which makes it possible to situate the scope of its inferences and assess its proximity to the university population of interest. Substantively, the findings indicated that most studies reported potential improvements in cognitive skills through feedback/personalization, although risks of dependence and reasoning quality persist. This antecedent is relevant to the perspective of feedback mechanisms because it demonstrates that the interpretation of the effect depends on the correspondence among construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

Study (4), published in 2025, provided evidence through system evaluation with real data. It used College Algebra Transactions and multiple internal studies, which makes it possible to situate the scope of its inferences and assess its proximity to the university population of interest. Substantively, the findings showed that diagnosis was close to or above 80% in several error

types, but an important fraction of hints was too general, incorrect, or revealed the answer. This antecedent is relevant to the perspective of feedback mechanisms because it demonstrates that the interpretation of the effect depends on the correspondence among construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

No, reference (5) provided evidence published in 2026 through a methodological equity study. The study used multiple scoring models and demographic subgroups, making it possible to situate the scope of its inferences and assess its proximity to the university population of interest. In substantive terms, the study showed that equity varied by demographic group and competence level; the measures distinguish predictive bias from real differences. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

Reference (6) provided evidence published in 2025 through a quasi-experimental pretest-posttest + acceptance design. The study had a sample of n=250 and included n=25 interviews, making it possible to situate the scope of its inferences and assess its proximity to the university population of interest. In substantive terms, the work showed that the intervention led to improvement in writing dimensions linked to critical thinking and to favorable acceptance of AI assistance. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

Reference (7) provided evidence published in 2025 through a systematic review of experimental studies. The study included 21 experimental studies, making it possible to situate the scope of its inferences and assess its proximity to the university population of interest. In substantive terms, the work showed that the effects are positive but heterogeneous; medium-duration interventions tend to show better results, and the effects on higher-order skills are variable. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

Reference (8) provided evidence published in 2025 through a PRISMA systematic review. The study included 39 studies (2023–2024), making it possible to situate the scope of its inferences and assess its proximity to the university population of interest. In substantive terms, the work showed that use can improve access and support, but overdependence is associated with risks to critical thinking, autonomy, and academic authenticity. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

Reference (9) provided evidence published in 2024 through a systematic review. The study included 57 articles, making it possible to situate the scope of its inferences and assess its proximity to the university population of interest. In substantive terms, the work identified support for learning and productivity along with problems of integrity, dependence, reliability, and the need for critical literacy. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

Reference (10) provided evidence published in 2025 using sequential mixed methods. The study worked with university students who were users of GPTutor, making it possible to situate the scope of its inferences and assess its proximity to the university population of interest. In substantive terms, the work showed that gPTutor favored engagement and personalized support; the quality of interaction and design of prompts/tutoring conditioned the experience. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as

an isolated demonstration sufficient to generalize to all educational scenarios.

Reference (11) provided evidence published in 2025 through an explanatory survey with moderation. The study used a sample of n=418 university students from a public university (China), making it possible to situate the scope of its inferences and assess its proximity to the target university population. In substantive terms, the work showed that dependence can elevate apparent self-efficacy but be negatively associated with achievement; teacher support modifies the relationship. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The work (12) was published in 2022 and provided evidence through system design and evaluation. The study involved users of ubiquitous environments, which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. Substantively, the work showed that adapting feedback to context and performance improved the relevance of the intervention and support for ubiquitous learning. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational settings.

The synthesized evidence (13) was published in 2021 and provided evidence via computational development and evaluation. The study used educational dialogue data, which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. Substantively, the work showed that deep discourse analysis made it possible to identify learning states and support feedback adapted to the student. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The examined contribution (14) was published in 2025 and provided evidence through a longitudinal qualitative case study. The study involved n=19 and lasted 13 weeks, which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. Substantively, the study showed that improvements in confidence and autonomy were reported; self-regulation was key to converting microcontents and GenAI into meaningful learning. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The study (15) was published in 2025 and provided evidence through a quasi-experimental and analytical design. The study used experimental and control groups in STEM, which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. Substantively, it showed that the experimental group outperformed the control group in programming and mathematics; greater interaction was related to progress, and 80% valued the adaptive feedback. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The study (16) was published in 2025 and provided evidence using a quantitative design with serial mediation plus a qualitative component. The study involved n=337 university students in China, which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. Substantively, the study showed that dependence on GenAI was related to lower independent, innovative, and critical thinking and to lower self-reflection; anxiety and expectations operated as mechanisms. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The work (17) was published in 2025 and provided evidence through a randomized controlled

trial. The study involved n=120, which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. Substantively, the work showed that the ChatGPT group exhibited lower long-term retention than traditional learning (57.5% vs 68.5% in the reported test), suggesting a risk of cognitive offloading. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The synthesized evidence (18) was published in 2025 and provided evidence through a conceptual and applied study. Substantively, the work proposes to redesign activities toward analysis, evaluation, and creation, avoiding delegating to GenAI the cognitive levels intended to be developed. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between construct, measurement procedure, and application context. Its contribution is considered in this review as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The examined contribution (19) provided evidence published in 2025 through a qualitative study of educational implementation. The study involved n=105 (72 undergraduate, 33 postgraduate), which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. In substantive terms, the study showed that the structured use of GenAI strengthened critical evaluation and responsible use; it also revealed risks of dependence and AI literacy gaps. This prior evidence is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between the construct, the measurement procedure, and the application context. In this review, its contribution is considered as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The study (20) provided evidence published in 2024 through a preregistered experiment with a control group. The study included a ChatGPT group (n=77) and a control group (n=68), which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. In substantive terms, the work showed that ChatGPT improved quality, elaboration, originality, and self-efficacy while reducing mental effort; risks of metacognitive calibration were observed. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between the construct, the measurement procedure, and the application context. In this review, its contribution is considered as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The research (21) provided evidence published in 2025 through a PRISMA systematic review. The study included 40 studies, which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. In substantive terms, the work showed that effectiveness depends on segmentation, timeliness, feedback, and alignment with objectives; it proposes a framework for integrating microlearning into broader experiences. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between the construct, the measurement procedure, and the application context. In this review, its contribution is considered as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

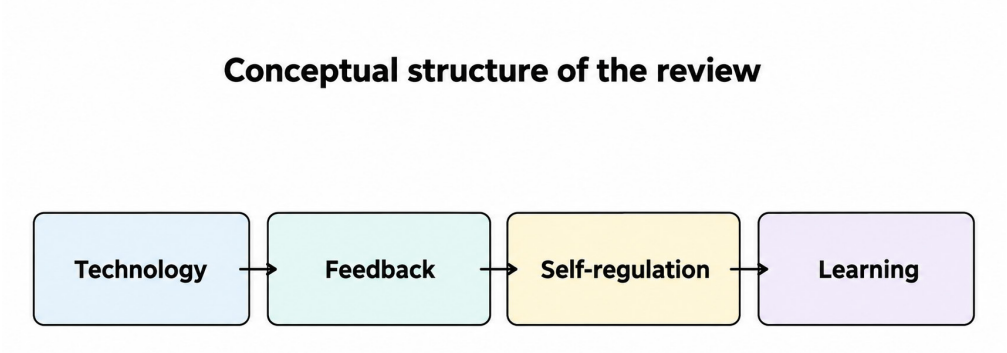
Hamida et al. (22) provided evidence published in 2026 through a technical accessibility assessment. The study involved students with special educational needs in higher education, which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. In substantive terms, the work identified that digital assessment accessibility requires compatibility with assistive technologies, autonomy, and integrity under exam restrictions. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between the construct, the measurement procedure, and the application context. In this review, its contribution is considered as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

The evidence synthesis (23) provided evidence published in 2025 through a bibliometric and systematic review. The study included 640 Scopus documents (2023–2025), which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. In substantive terms, the work showed that the literature converges on AI literacy,

information verification, and design of tasks that require justifying, contrasting, and reflecting on AI outputs. This precedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between the construct, the measurement procedure, and the application context. In this review, its contribution is considered as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

Agustín et al. (24) provided evidence published in 2024 through scale development and validation. The study involved students from a Venezuelan university, which allows the scope of its inferences to be situated and its proximity to the university population of interest to be assessed. In substantive terms, the work showed that a model of 22 variables in four dimensions with satisfactory psychometric properties was obtained. This antecedent is relevant to the perspective of feedback mechanisms because it shows that the interpretation of the effect depends on the correspondence between the construct, the measurement procedure, and the application context. In this review, its contribution is considered as a piece of the comparative pattern, not as an isolated demonstration sufficient to generalize to all educational scenarios.

**Figure 1. Conceptual structure of the review**




**Note. Prepared by the authors based on the analytical axes defined for the manuscript.**

The figure organizes the reasoning of the article as a conceptual sequence and not as a universal template. The components represented correspond to the specific problem of this review and show the transition from the inputs or initial conditions to the educational interpretation. Its usefulness lies in making visible the relationship between constructs that, in individual studies, usually appear fragmented across methodological sections and results. This representation guides the subsequent synthesis and facilitates distinguishing the elements that act as conditions, mechanisms, or results within the corpus.

## Review questions

Question 1. What patterns of evidence enable an understanding of dimension 1 of feedback mechanisms in relation to the effectiveness of generative artificial intelligence in formative assessment and learning feedback in university students, and what methodological or contextual conditions modify their interpretation?

Question 2. What patterns of evidence enable an understanding of dimension 2 of feedback mechanisms in relation to the effectiveness of generative artificial intelligence in formative assessment and learning feedback in university students, and what methodological or contextual conditions modify their interpretation?

Question 3. What patterns of evidence enable an understanding of dimension 3 of feedback mechanisms in relation to the effectiveness of generative artificial intelligence in formative assessment and learning feedback in university students, and what methodological or contextual conditions modify their interpretation?

Question 4. What patterns of evidence enable an understanding of dimension 4 of feedback mechanisms in relation to the effectiveness of generative artificial intelligence in formative assessment and learning feedback in university students, and what methodological or contextual conditions modify their interpretation?

## METHODS

### PRISMA Method

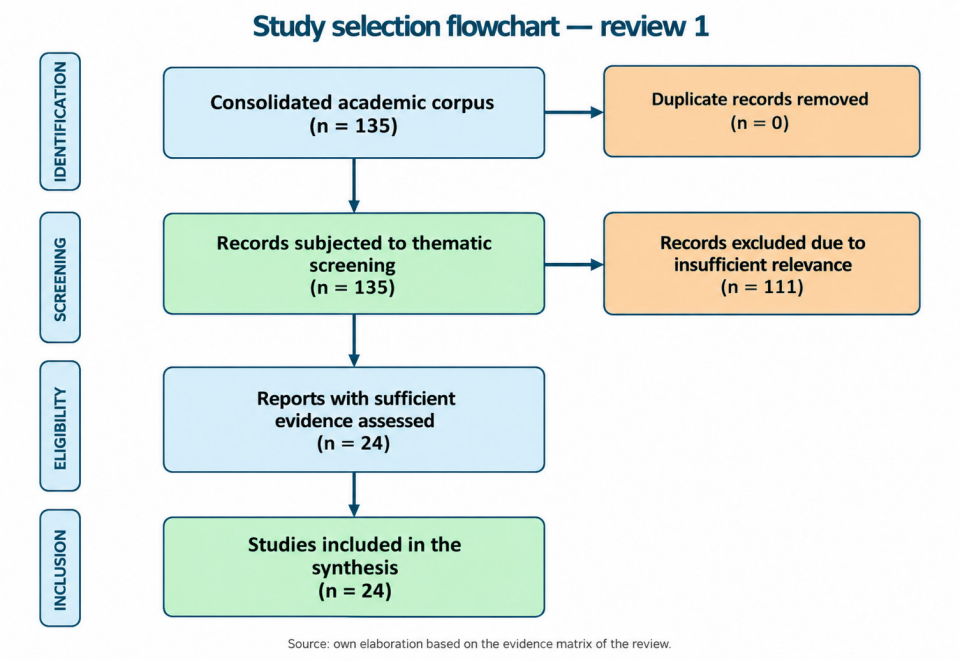
A systematic review was developed to synthesize pertinent academic evidence on the topic of the manuscript. The unit of analysis was the academic study or reference with a traceable source, identifiable methodology, and retrievable results. The consolidated working academic corpus consisted of 135 records previously gathered in the evidence matrices. A thematic and documentary selection was applied to this set to establish the final corpus. The procedure was organized according to the PRISMA logic of identification, screening, eligibility, and inclusion, adapted to the consolidated nature of the provided sources.

The inclusion criteria required a direct relationship with the effectiveness of generative artificial intelligence in formative assessment and learning feedback in university students, correspondence with higher education or with instruments applicable to that level, sufficient description of the method, and availability of interpretable results. A total of 111 records were excluded because of lower relevance to the specific question, because they addressed peripheral populations or outcomes, or because they did not provide sufficient methodological evidence for the focused synthesis. The final corpus consisted of 24 studies. The selection was carried out on the basis of the information documented in the matrices, and attributing counts not supported by the files to any particular bibliographic database was avoided.

Data extraction was structured using a matrix containing authors, year, title, source, DOI or traceable link, origin, study type, sample or corpus, summarized methodology, main results, level of evidence, and relevance. This organization allowed comparison of works with heterogeneous designs without reducing them to a single metric. The synthesis combined descriptive characterization of the corpus and thematic analysis of the results, attending to convergences, contradictions, implementation conditions, and limits of generalization.

Evidence evaluation focused on whether the retrieved information was sufficient to answer the review question. The clarity of the design, the explicitness of the sample or corpus, the correspondence between method and results, and the traceability of the source were considered jointly. Meta-analysis was not performed when heterogeneity of outcomes, instruments, populations, or metrics prevented a responsible quantitative combination. In such cases, a comparative synthesis was prioritized, preserving differences between studies and avoiding transforming non-equivalent results into a single estimator.

**Figure 2. Flow of study selection according to the PRISMA workflow**




Note. The flow represents 135 normalized academic records, with no duplicates in the consolidated corpus; 111 records did not meet the thematic relevance threshold, and 24 studies were included in the final synthesis.

To strengthen transparency, the results were organized into characterization and synthesis tables, while the figures represent the selection flow and the temporal distribution of the corpus. Each table was constructed from the fields of the provided matrix. Subsequent interpretations were developed based on observable patterns in the set and not on additional assumptions. This criterion is particularly important in educational reviews, where variation across disciplines, technologies, instruments, and institutional contexts can substantially modify the magnitude and meaning of the findings.

The flow documents the deduplication and selection of evidence. The 135 records in the consolidated corpus were normalized by DOI or title, so no duplicates were identified at this stage. The thematic screening excluded 111 records due to insufficient correspondence with the specific review question, while 24 studies had sufficiently recoverable methodological evidence and results to be integrated into the synthesis. The sequence permits distinguishing document identification from eligibility judgment and prevents interpreting the reduction of the corpus as an arbitrary loss of information.

## RESULTS

### Evidence Mapping

The final corpus comprised 24 studies published during 2021–2026. The temporal distribution reveals an active and heterogeneous research field, with studies responding to different stages of maturation of the problem. Some focus on demonstrating feasibility or association, while others advance toward validation, comparison of strategies, or analysis of implementation conditions. This temporal breadth allows us to observe not only favorable results, but also shifts in how the quality of assessment is conceptualized and in the type of evidence considered necessary to support educational decisions.

The methodological characterization confirms that the effectiveness of generative artificial intelligence in formative assessment and learning feedback in university students does not constitute a homogeneous object. The set includes designs with different purposes and varying levels of control over study conditions. This diversity broadens the scope of the review, although it also requires interpreting the findings according to the function of each design. Intervention studies report on observed changes under defined conditions; instrumental studies provide evidence on measurement; observational analyses describe associations; and previous reviews allow recognition of field trends without replacing the evaluation of primary studies.

The samples and corpora described in the studies show considerable variability. This variation is relevant because the transferability of the results depends on the size, composition, and context of the participants, as well as on the academic discipline and the technological environment. Overall, the evidence suggests that the most interpretable results are those that specify the population, the evaluative task, and the mechanism by which the intervention or instrument should produce the expected effect. When any of these elements remains poorly defined, the usefulness of the finding for institutional decisions decreases.

The synthesis of results reveals a central pattern linked to feedback mechanisms: favorable effects or properties do not appear as automatic attributes of the technology or instrument, but as outcomes conditioned by design, implementation, data quality, and pedagogical alignment. The studies converge on the importance of articulating the tool with clear learning objectives, understandable evaluation criteria, and monitoring mechanisms. Divergences are concentrated in the magnitude of the effects, stability across contexts, and the capacity to sustain outcomes when institutional conditions change.

The table shows the methodological diversity of the corpus and allows observing that the studies do not offer the same type of inference. The presence of designs with distinct objectives requires interpreting each result according to the question that the study can answer. This differentiation avoids directly comparing indicators that fulfill different functions and helps identify where real convergence exists. It also shows that the robustness of a conclusion depends on the combination of methodological clarity, sample description, and sufficiency of the recovered evidence.

The temporal distribution allows us to situate the recent intensity of scientific production and to recognize whether the field is undergoing expansion, consolidation, or methodological transition. Rather than interpreting the number of publications as a measure of quality, the figure serves to contextualize the changes in questions, technologies, and instruments observed in the

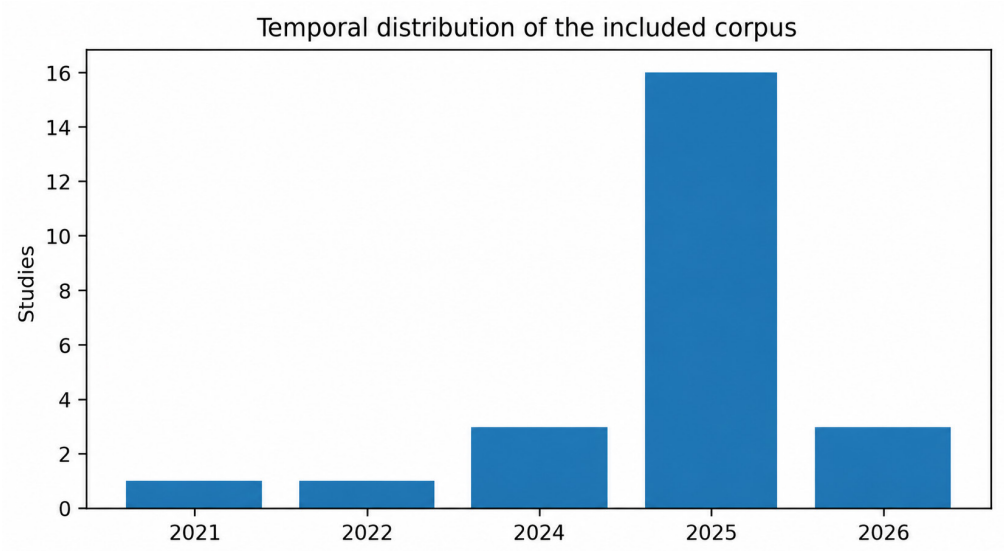
corpus. The joint reading with the methodological characterization makes it possible to identify whether the most recent works expand the evidence through more robust designs or whether exploratory approaches continue to predominate.

**Table 1. Methodological characterization of the included corpus.**

| Year | Design/methodology | Sample or Corpus | Evidence |
|---|---|---|---|
| 2025 | Exploratory mixed methods | Survey n=40; scripts n=6; interviews n=2 | Mean |
| 2026 | Empirical comparison of models | Analyses with gender, ELL, and special education subgroups | High |
| 2025 | Systematic literature review | 158 studies (2021–2024) | Mean |
| 2025 | System evaluation with real data | College Algebra transactions; multiple internal studies | Mean |
| 2026 | Methodological study of equity | Multiple scoring models and demographic subgroups | High |
| 2025 | Quasi-experimental pretest-posttest + acceptance | n=250; interviews n=25 | Mean |
| 2025 | Systematic review of experimental studies | 21 experimental studies | Mean |
| 2025 | PRISMA systematic review | 39 studies (2023–2024) | Mean |
| 2024 | Systematic review | 57 articles | Mean |
| 2025 | Sequential mixed methods | University student users of GPTutor | Mean |
| 2025 | Explanatory survey with moderation | n=418 university students from a public university (China) | Mean |
| 2022 | System design/evaluation | Ubiquitous environment users | Mean |

Note. Data extracted from the evidence matrix provided for this review.

**Figure 3. Temporal distribution of included studies**




Note. Frequencies calculated from the publication year recorded in the evidence matrix.

## Patterns of effectiveness and conditions of use

Another cross-cutting pattern is the relevance of evidence quality. Studies with explicit procedures make it possible to better distinguish among real improvement, favorable perception, and mere ease of use. In contrast, when the result is supported mainly by self-report or short-term indicators, interpretation requires greater caution. The review shows that the practical utility of a strategy depends not only on whether it produces a statistical difference or a positive evaluation, but on whether the measurement is consistent with the construct and whether the intervention can be sustainably integrated into real educational processes.

In the subset of evidence corresponding to this axis, results were observed that, considered jointly, can be summarized as follows: Collaborative and reflective use favored phases of cognitive presence; passive AI-directed use showed less critical elaboration.

Adjusted GPT-4o achieved the highest accuracy and overall fairness among the models compared. Most reported potential improvements in cognitive skills through feedback/personalization, although risks of dependence and quality of reasoning persist. The comparative reading indicates that these findings should not be interpreted as equivalent, because they come from different designs and metrics. However, their thematic convergence allows identifying recurrent conditions and formulating an overall interpretation focused on the relationship between quality of design, student response, and usefulness of the information produced to guide educational decisions.

The results observed in the subset of evidence corresponding to this axis, considered jointly, can be summarized as follows: Diagnosis was close to or above 80% for several error types, but a significant fraction of hints were too general, incorrect, or revealed the answer. Fairness varied by demographic group and proficiency level; the measures distinguished predictive bias from real differences. Reports indicated improvement in writing dimensions linked to critical thinking and favorable acceptance of AI assistance. The comparative reading indicates that these findings should not be interpreted as equivalent, because they derive from different designs and metrics. However, their thematic convergence allows the identification of recurrent conditions and the formulation of an overall interpretation focused on the relationship among design quality, student response, and the usefulness of the information produced to guide educational decisions.

The results observed in the subset of evidence corresponding to this axis, considered jointly, can be summarized as follows: The effects are positive but heterogeneous; medium-duration interventions tend to show better results, and the effects on higher-order skills are variable. Use can improve access and support, but overdependence is associated with risks to critical thinking, autonomy, and academic authenticity. Support for learning and productivity is identified, along with problems of integrity, dependence, reliability, and the need for critical literacy. The comparative reading indicates that these findings should not be interpreted as equivalent, because they derive from different designs and metrics. However, their thematic convergence allows identifying recurrent conditions and formulating an overall interpretation focused on the relationship among design quality, student response, and the usefulness of the information produced to guide educational decisions.

The results observed in the subset of evidence corresponding to this axis, considered together, can be summarized as follows: GPTutor favored engagement and personalized support; the quality of interaction and prompt/tutoring design conditioned the experience. Dependence can raise apparent self-efficacy but be negatively associated with achievement; teacher support modifies the relationship. Adaptation of feedback to context and performance improved the relevance of the intervention and support for ubiquitous learning. The comparative reading indicates that these findings should not be interpreted as equivalent, because they come from different designs and metrics. However, their thematic convergence allows identifying recurrent conditions and formulating an overall interpretation centered on the relationship among design quality, student response, and the usefulness of the information produced to guide educational decisions.

The results observed in the subset of evidence corresponding to this axis, considered jointly, can be summarized as follows: The in-depth discourse analysis allowed identifying learning states and supporting feedback adapted to the student. Improvements in confidence and autonomy were reported; self-regulation was key to turning microcontent and GenAI into meaningful learning. The experimental group outperformed the control group in programming and mathematics; greater interaction was related to progress, and 80% valued adaptive feedback. The comparative reading indicates that these findings should not be interpreted as equivalent, because they come from different designs and metrics. However, their thematic convergence allows identifying recurrent conditions and formulating an overall interpretation focused on the relationship among quality of design, student response, and usefulness of the information produced to guide educational decisions.

The results observed in the subset of evidence corresponding to this axis, considered jointly, can be summarized as follows: GenAI dependence was related to less independent/innovative/critical thinking and less self-reflection; anxiety and expectations operated as mechanisms. The ChatGPT group showed lower long-term retention than traditional learning (57.5% vs 68.5% in the reported test), suggesting a risk of cognitive offloading. Redesigning activities toward analysis, evaluation, and creation is proposed, avoiding delegating to GenAI the cognitive levels that are intended to be developed. The comparative reading indicates that these findings should not be interpreted as equivalent, because they come from different designs and metrics. However, their thematic convergence allows identifying recurrent conditions and formulating an overall interpretation centered on the relationship among design quality, student response, and the

usefulness of the information produced to guide educational decisions.

The results observed in the subset of evidence corresponding to this axis, considered together, can be summarized as follows: The structured use of GenAI strengthened critical evaluation and responsible use; it also evidenced risks of dependence and gaps in AI literacy. ChatGPT improved quality, elaboration, originality, and self-efficacy and reduced mental effort; risks of metacognitive calibration were observed. Effectiveness depends on segmentation, timeliness, feedback, and alignment with objectives; a framework is proposed to integrate microlearning into broader experiences. The comparative reading indicates that these findings should not be interpreted as equivalent, because they come from different designs and metrics. However, their thematic convergence makes it possible to identify recurrent conditions and formulate an overall interpretation centered on the relationship among design quality, student response, and the usefulness of the information produced for guiding educational decisions.

In the subset of evidence corresponding to this axis, the results observed, considered together, can be summarized as follows: they identify that digital assessment accessibility requires compatibility with assistive technologies, autonomy, and integrity under examination restrictions. The literature converges on IA literacy, information verification, and task design that compels justification, contrast, and reflection on IA outputs. A model of 22 variables in 4 dimensions with satisfactory psychometric properties was obtained. The comparative reading indicates that these findings should not be interpreted as equivalent, because they stem from different designs and metrics. However, their thematic convergence allows identifying recurrent conditions and formulating an overall interpretation centered on the relationship between design quality, student response, and the usefulness of the information produced to guide educational decisions.

**Table 2. Synthesis of relevant findings**

| Study | Recovered finding | Relevance |
|---|---|---|
| Exploring the Impact of Generative AI ChatGPT on Critical Thinking in Higher Education: Pa | Collaborative and reflective use favored phases of cognitive presence; passive AI-directed use showed less critical elaboration. | Included because of direct relevance to the |
| Accuracy and fairness of generative AI in automated essay scoring: Comparing GPT-4o, featu | Fine-tuned GPT-4o achieved the highest overall accuracy and fairness among the compared models. | Included because of direct relevance to the |
| Can Generative AI Revolutionise Academic Skills Development in Higher Education? A Systema | The majority reported potential improvements in cognitive skills through feedback and personalization, although risks of dependency and quality persist. | Included because of direct relevance to the |
| Generating In-Context, Personalized Feedback for Intelligent Tutors with Large Language Mo | Diagnosis was close to/above 80% in several error types, but a significant fraction of clues was too general, incorrect, or revealed the res | Included because of direct relevance to the |
| Assessing fairness in AI-assisted writing scoring: Developing fairness measures to detect | Fairness varied by demographic group and level of competence; the measures distinguish predictive bias from real differences. | Included because of direct relevance to the |
| An AI-assisted critical thinking intervention to enhance undergraduate EFL learners' writi | The intervention reported improvements in writing dimensions associated with critical thinking and favorable acceptance of AI assistance. | Included because of direct relevance to the |
| ChatGPT Interventions in Higher Education: A Systematic Review of Experimental Studies | The effects are positive but heterogeneous; interventions of medium duration tend to show better results and the effects on skills sup | Included because of direct relevance to the |
| Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing | The use can improve access and support, but overdependence is associated with risks to critical thinking, autonomy, and academic authenticity. | Included because of direct relevance to the |
| ChatGPT in higher education: A systematic literature review and research challenges | Support for learning and productivity is identified, along with issues of integrity, dependence, reliability, and the need for critical literacy. | Included because of direct relevance to the |
| Promoting student engagement with GPTutor: An intelligent tutoring system powered by gener | GPTutor promoted engagement and personalized support; the quality of interaction and the design of prompts and tutoring conditioned the experience. | Included because of direct relevance to the |

Note. Textual synthesis based exclusively on the results recorded in the matrix.

The synthesis shows that the favorable evidence is distributed across different types of outcomes and that the relevance of each study depends on its proximity to the central construct. The findings are not interpreted as interchangeabe: some describe performance, others measurement properties, student experience, accessibility, or model performance. The comparison allows recognizing convergences without losing the specificity of each result, and it constitutes the basis for the discussion of mechanisms and implementation conditions.

## DISCUSSION

The synthesis supports the view that the problem of the effectiveness of generative artificial intelligence in formative assessment and learning feedback in university students must be interpreted from a mechanism-based feedback logic, rather than from a simplified opposition between effective and ineffective technology. The reviewed studies show that the performance of a strategy depends on the interaction among assessment purpose, student characteristics, task design, measurement quality, and institutional capacity to use the generated information. This conclusion shifts the focus from the isolated tool to the assessment system in which it is embedded.

An important consequence is that technological innovation does not eliminate the classical problems of validity, reliability, fairness, and utility. On the contrary, it can make them more visible by increasing the scale and speed of decisions. The reviewed evidence suggests that the best implementations are those that make criteria explicit, preserve feedback opportunities, and maintain human review mechanisms when the consequences of a decision are significant. In this sense, digitalization should be understood as a transformation of the evaluative process, not as an automatic replacement of academic judgment.

The observed heterogeneity also explains why it is not appropriate to reduce the field to a single effect size. Studies differ in constructs, duration, instruments, disciplines, samples, and metrics. This variation is not merely a limitation; it also provides information about the contexts in which a strategy works and about the conditions that must be considered before transferring it. Comparative synthesis allows the recognition of patterns of consistency without erasing differences that are essential for educational decision-making.

From a methodological standpoint, the corpus shows a transition from studies focused on adoption and perception toward designs that seek to demonstrate quality of measurement, impact on outcomes, and sustainability of the intervention. However, gaps persist related to longitudinal comparisons, replication across institutions, subgroup analyses, and documentation of unintended effects. The future agenda should prioritize designs that connect learning outcomes with implementation processes and that inform not only whether an intervention works, but for whom, under what conditions, and with what pedagogical or organizational costs.

The main contribution of this review is to organize the evidence so that the findings can be used for academic decisions without overextending their scope. Convergence across studies provides arguments for adopting promising practices, but methodological diversity requires maintaining a logic of continuous evaluation. Institutions should treat each implementation as an intervention amenable to monitoring, with indicators defined before its deployment and mechanisms to review outcomes, biases, student experience, and coherence with curricular objectives.

The research (1) offers a useful point of contrast for this interpretation because it documents that collaborative and reflective use favored phases of cognitive presence, whereas passive AI-directed use showed less critical elaboration. This finding acquires greater significance when compared with the rest of the corpus: it does not act as self-sufficient evidence, but rather as an observation that helps to specify mechanisms, limits, and conditions of application. Systematic reading thus makes it possible to avoid 2 frequent extremes: excessive generalization based on a single study, and the loss of contextual information when heterogeneous results are synthesized.

No. (2) offers a useful point of contrast for this interpretation because it documents that fine-tuned gPT-4o achieved the highest accuracy and overall fairness among the compared models. This finding acquires greater significance when compared with the rest of the corpus: it does not act as self-sufficient evidence, but rather as an observation that helps to specify mechanisms, limits, and conditions of application. The systematic reading thus makes it possible to avoid 2 frequent extremes: excessive generalization from a single study, and the loss of contextual information when synthesizing heterogeneous results.

The synthesized evidence (3) offers a useful point of contrast for this interpretation because the study documents that the majority reported potential improvements in cognitive skills through feedback/personalization, although risks of dependence and reasoning quality persist. This finding acquires greater significance when compared with the rest of the corpus: it does not act as self-sufficient evidence, but rather as an observation that helps to specify mechanisms, limits, and conditions of application. Systematic reading thus makes it possible to avoid two frequent extremes: excessive generalization from a single study and the loss of contextual information when heterogeneous results are synthesized.

The examined contribution (4) offers a useful point of contrast for this interpretation because the study documents diagnosis close to or above 80% in various error types, but an important

fraction of cues was too general, incorrect, or revealed the answer. This finding acquires greater significance when compared with the rest of the corpus: it does not act as self-sufficient evidence, but rather as an observation that helps to specify mechanisms, limits, and conditions of application. Systematic reading thus makes it possible to avoid two frequent extremes: excessive generalization from a single study and the loss of contextual information when heterogeneous results are synthesized.

The study (5) offers a useful point of contrast for this interpretation because it documents that fairness varied by demographic group and competency level; the measures distinguish predictive bias from real differences. This finding acquires greater significance when compared with the rest of the corpus: it does not act as self-sufficient evidence, but rather as an observation that helps to specify mechanisms, limits, and conditions of application. Systematic reading thus makes it possible to avoid two frequent extremes: excessive generalization from a single study and the loss of contextual information when heterogeneous results are synthesized.

Research (6) offers a useful point of contrast for this interpretation because the study documents the intervention's reported improvement in writing dimensions linked to critical thinking and a favorable acceptance of AI assistance. This finding acquires greater significance when compared with the rest of the corpus: it does not act as self-sufficient evidence, but rather as an observation that helps to specify mechanisms, limits, and conditions of application. Systematic reading thus makes it possible to avoid two frequent extremes: excessive generalization from a single study and the loss of contextual information when heterogeneous results are synthesized.

## Implications for University Evaluation

The practical implications derived from the review should translate into gradual decisions. For generative artificial intelligence to be effective in formative assessment and learning feedback in university students, responsible implementation requires defining in advance the construct to be measured, establishing success indicators, identifying potentially affected groups, and planning review mechanisms. The available evidence favors strategies that integrate innovation with teacher accompaniment and continuous evaluation of results.

At the institutional level, adoption should be accompanied by data governance protocols, transparency criteria, and communication mechanisms for students and teachers. These conditions increase the interpretability of the results and reduce the risk of using indicators outside the context for which they were designed. The evaluation of a tool should consider both technical performance and pedagogical and distributive consequences.

For research, the next step is to better connect efficacy studies with implementation studies. It is necessary to know not only whether a strategy produces changes, but also what components explain those changes, what resources it demands, and how it behaves in diverse populations. This orientation would allow evidence to be accumulated more coherently and reduce the fragmentation that currently characterizes part of the literature.

## Limitations

The main limitation of the corpus is its heterogeneity. Differences between designs, instruments, populations, and outcomes restrict the possibility of producing a common estimator and make it necessary to interpret patterns from a comparative logic. This limitation does not invalidate the synthesis, but it does condition the scope of the conclusions and reinforces the need for replications with comparable protocols.

A concentration of studies is also observed in certain contexts and technologies, which reduces certainty about transferability to institutions with different infrastructure, student profiles, or assessment cultures. Future research should document with greater precision the implementation conditions, participant characteristics, and decisions made during the deployment of the tools.

Finally, the review is based on the consolidated corpus documented in the provided matrices. The conclusions faithfully represent that set and do not extend to records that are not part of the available evidence. This delimitation preserves traceability and allows future updates to incorporate new studies without retrospectively altering the analytical basis used in the present manuscript.

## Conclusions

This systematic review shows that the effectiveness of generative artificial intelligence in formative assessment and learning feedback among university students constitutes a field with

sufficient evidence to identify patterns, although it remains heterogeneous in designs, populations, and metrics. The central finding is that the quality of decisions depends less on the isolated presence of a technology or instrument than on its integration with learning objectives, explicit criteria, and monitoring procedures. The feedback mechanisms perspective makes it possible to understand why similar interventions can produce different results and why the evaluation of effectiveness must simultaneously consider outcomes, conditions, and measurement quality.

In applied terms, higher education institutions should adopt a gradual implementation approach, with controlled testing, documentation of results, and periodic review of effects on different student groups. The evidence supports the use of digital strategies when there is pedagogical alignment and when the information generated is used to improve teaching and learning. It also indicates the need to avoid automatic decisions based on single indicators, especially when academic consequences are significant or when there are differences in access, technological experience, or support needs.

The research agenda should move toward longitudinal studies, multicenter comparisons, and designs that more precisely document mechanisms and differential effects. It is also necessary to improve the comparability of indicators and transparency in describing instruments, algorithms, and intervention procedures. These priorities will make it possible to transform a field characterized by rapid innovation into a more cumulative and replicable evidence base that is useful for university policies. The review provides a structure for this advancement by distinguishing consistent findings, implementation conditions, and gaps that still require research.

## REFERENCES

1. Nesma Ragab Nasr; Chih-Hsiung Tu; Jennifer Werner; Tonia Bauer; et al. Exploring the Impact of Generative AI ChatGPT on Critical Thinking in Higher Education: Passive AI-Directed Use or Human–AI Supported Collaboration? Education Sciences. 2025;15(9):1198. https://doi.org/10.3390/educsci15091198
2. Accuracy and fairness of generative AI in automated essay scoring: Comparing GPT-4o, feature-based models, and human raters. Assessing Writing. 2026;69:101047. https://doi.org/10.1016/j.asw.2026.101047
3. Kangwa Daniel; M. M. Msambwa; Z. Wen. Can Generative AI Revolutionise Academic Skills Development in Higher Education? A Systematic Literature Review. European Journal of Education. 2025;60(1):e70036. https://doi.org/10.1111/ejed.70036
4. Jennifer M. Reddig; Arav Arora; Christopher J. MacLellan. Generating In-Context, Personalized Feedback for Intelligent Tutors with Large Language Models. International Journal of Artificial Intelligence in Education. 2025;35:3459–3500. https://doi.org/10.1007/s40593-025-00505-6
5. Assessing fairness in AI-assisted writing scoring: Developing fairness measures to detect predictive bias in automated essay scoring. Assessing Writing. 2026. https://doi.org/10.1016/j.asw.2026.101066
6. Xue Yin; Kun Dou; Jinlong Han. An AI-assisted critical thinking intervention to enhance undergraduate EFL learners' writing proficiency. Studies in Educational Evaluation. 2025;86:101480. https://doi.org/10.1016/j.stueduc.2025.101480
7. Yiwen Jin; Lies Sercu. ChatGPT Interventions in Higher Education: A Systematic Review of Experimental Studies. Journal of Computer Assisted Learning. 2025;41(4):e70072. https://doi.org/10.1111/jcal.70072
8. Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration. Social Sciences & Humanities Open. 2025. https://doi.org/10.1016/j.ssaho.2025.101866
9. Maria Ijaz Baig; Elaheh Yadegaridehkordi. ChatGPT in the higher education: A systematic literature review and research challenges. International Journal of Educational Research. 2024;127:102411. https://doi.org/10.1016/j.ijer.2024.102411
10. Haoran Bai; Wing Cheung Lui; Paul Vinod Khiatani. Promoting student engagement with GPTutor: An intelligent tutoring system powered by generative AI. International Journal of Educational Technology in Higher Education. 2025;22:77. https://doi.org/10.1186/s41239-025-00571-9
11. The paradox of self-efficacy and technological dependence: Unraveling generative AI's impact on university students' task completion. The Internet and Higher Education. 2025;65:100978. https://doi.org/10.1016/j.iheduc.2024.100978
12. Marta S. Tabares; Paola Vallejo; Alex Montoya; Daniel Correa. A feedback model applied in a ubiquitous microlearning environment using SECA rules. Journal of Computing in Higher Education. 2022;34:462–488. https://doi.org/10.1007/s12528-021-09306-x
13. Matt Grenander; Robert Belfer; Ekaterina Kochmar; Iulian V. Serban; François St-Hilaire; Jackie C. K. Cheung. Deep Discourse Analysis for Generating Personalized Feedback in Intelligent Tutor Systems. Proceedings of the AAAI Conference on Artificial Intelligence. 2021;35(17):15534–15544. https://doi.org/10.1609/aaai.v35i17.17829
14. Lucas Kohnke; Di Zou; Haoran Xie. Microlearning and generative AI for pre-service teacher education: a qualitative case study. Education and Information Technologies. 2025;30(15):21221–21248. https://doi.org/10.1007/s10639-025-13606-5
15. William Villegas-Ch; Diego Buenano-Fernandez; Alexandra Maldonado Navarro; Aracely Mera-Navarrete. Adaptive intelligent tutoring systems for STEM education: analysis of the learning impact and effectiveness of personalized feedback. Smart Learning Environments. 2025;12:41. https://doi.org/10.1186/s40561-025-00389-y
16. Ting Huang; Chenze Wu. The chain mediating effect of academic anxiety and performance expectations between academic self-efficacy and generative AI reliance. Computers and Education Open. 2025;9:100275. https://doi.org/10.1016/j.caeo.2025.100275
17. André Barcaui. ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. Social Sciences & Humanities Open. 2025;12:102287. https://doi.org/10.1016/j.ssaho.2025.102287
18. Chahna Gonsalves. Generative AI's Impact on Critical Thinking: Revisiting Bloom's Taxonomy. Journal of Marketing Education. 2026;48(1). https://doi.org/10.1177/02734753241305980
19. Mahmoud Elkhodr; Ergun Gide. The SAGE framework for developing critical thinking and responsible generative AI use in cybersecurity education. Discover Education. 2025;4:517. https://doi.org/10.1007/s44217-025-00935-3
20. Marek Urban; Filip Děchtěrenko; Jiří Lukavský; Veronika Hrabalová; Filip Svacha; Cyril Brom; Kamila Urban. ChatGPT improves creative problem-solving performance in university students: An experimental study. Computers & Education. 2024;215:105031. https://doi.org/10.1016/j.compedu.2024.105031
21. Wali Khan Monib; Atika Qazi; Rosyzie Anna Apong. Microlearning beyond boundaries: A systematic review and a novel framework for improving learning outcomes. Heliyon. 2025;11(2):e41413. https://doi.org/10.1016/j.heliyon.2024.e41413
22. Hamida Al-Maamari; Manjit Singh Sidhu; Shaik Mazhar Hussain; Abbas M. Al-Ghaili; Kirandeep Kaur Sidhu. Accessible ICT-Enabled Examination Technologies for Students with Special Educational Needs in Higher Education: A Technical Evaluation of Usability, Reliability, Accommodation Quality, and Assessment Fairness. International Journal of Special Education. 2026;41(17s):635–654.
23. Achmad Salido; Irman Syarif; Melyani Sari Sitepu; Suparjan; Prima Rias Wana; Ryan Taufika; Rahyuni Melisa. Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies. Social Sciences & Humanities Open. 2025;12:101924. https://doi.org/10.1016/j.ssaho.2025.101924

24. Agustín Mejías-Acosta; Mayra D'Armas Regnault; Eduardo Vargas-Cano; Jesennia Cárdenas-Cobo; Cristian Vidal-Silva. Assessment of digital competencies in higher education students: development and validation of a measurement scale. Frontiers in Education. 2024. https://doi.org/10.3389/feduc.2024.1497376

## FUNDING

None.

## CONFLICT OF INTEREST

None exist.

## AUTHORSHIP CONTRIBUTIONS

Conceptualization: Luis David Bastidas González.
Data curation: Luis David Bastidas González.
Formal analysis: Luis David Bastidas González.
Research: Luis David Bastidas González.
Methodology: Luis David Bastidas González.
Supervision: Luis David Bastidas González.
Validation: Luis David Bastidas González.
Visualization: Luis David Bastidas González.
Writing – original draft: Luis David Bastidas González.
Writing – review and editing: Luis David Bastidas González.