Online Journal of Communication and Media Technologies, 2026, **16**(2), e202634
e-ISSN: 1986-3497

OPEN ACCESS

Research Article

Does generative artificial intelligence transform university learning? A correlational meta-analysis of educational outcomes

Alejandro Valencia-Arias [1*]
0000-0001-9434-6923

Erick Oswaldo Salazar Montoya [2]
0000-0001-5152-8997

Lelis Grabiel Palacios Silva [3]
0009-0001-9399-4395

Gustavo Adolfo Moreno López [4]
0000-0002-4094-0940

Sebastián Arias García [5]
0009-0001-3515-4072

Paula Andrea Rodríguez-Correa [5]
0000-0002-9748-0148

Wilmer Londoño-Celis [6]
0000-0003-1399-2654

[1] Escuela de Ingeniería Industrial, Universidad Señor de Sipán, Chiclayo, PERU
[2] Ciencias Económicas, Universidad Señor de Sipán, Chiclayo, PERU
[3] Ciencias de la Salud, Universidad Señor de Sipán, Chiclayo, PERU
[4] Institución Universitaria Marco Fidel Suárez, Bello, COLOMBIA
[5] Facultad de Ciencias Económicas y Administrativas, Instituto Tecnológico Metropolitano, Antioquia, COLOMBIA
[6] Facultad de Humanidades y Ciencias Sociales, Corporación Universitaria Americana, Medellín, COLOMBIA
[*] Corresponding author: valenciajho@uss.edu.pe

Citation: Valencia-Arias, A., Salazar Montoya, E. O., Palacios Silva, L. G., Moreno López, G. A., Arias García, S., Rodríguez-Correa, P. A., & Londoño-Celis, W. (2026). Does generative artificial intelligence transform university learning? A correlational meta-analysis of educational outcomes. *Online Journal of Communication and Media Technologies, 16*(2), Article e202634. https://doi.org/10.30935/ojcmt/18593

ARTICLE INFO

Received: 23 Feb 2026

Accepted: 2 May 2026

ABSTRACT

Generative artificial intelligence (GenAI) refers to systems capable of producing original content from large volumes of data and deep learning models. There has been an increasing adoption of GenAI in higher education as cognitive support, a tool for academic production, and a resource for personalized learning. However, the extant empirical evidence demonstrates a lack of consensus and, in certain instances, a paucity of concordance, impeding a holistic comprehension of its educational impact. In this context, the objective of this research is to estimate the overall relationship between GenAI and university educational outcomes through a correlational meta-analysis. The study employs a quantitative approach and is conducted in accordance with the PRISMA 2020 guidelines as an international reporting standard. This ensures transparency, reproducibility, and rigor in the identification, selection, evaluation, and synthesis of evidence. The findings confirm that GenAI constitutes a relevant educational phenomenon whose impact cannot be interpreted from simplistic or deterministic perspectives. Whilst the relationship with educational outcomes is consistent and methodologically stable, it is strongly influenced by context, pedagogical decisions and usage patterns. The absence of homogeneous patterns that would permit direct generalizations is a salient finding, underscoring the necessity for nuanced and critical analyses. This demonstrates that the educational transformation associated with GenAI is contingent not on the technology itself, but rather on its pedagogical and institutional integration.

Keywords: generative artificial intelligence, university learning, educational outcomes, correlational meta-analysis, higher education

INTRODUCTION

Generative artificial intelligence (GenAI) refers to computer systems capable of generating original content, such as text, images, or code, from the processing of large volumes of data using deep learning models. There is a growing presence of GenAI in higher education (Beauvais et al., 2014; Iqbal et al., 2021). Within the domain of higher education, GenAI is employed as a cognitive support instrument, a resource for the development of academic content, an automated feedback mechanism, and a medium for personalized learning. This finding aligns with the findings of empirical studies that have identified a strong correlation between technological, pedagogical, and contextual factors and academic performance (Ermold, 2011; Haider & Hussain, 2014; Ishak et al., 2023). Its incorporation is framed within the processes of digital transformation and the consolidation of data-driven educational models, with the potential to improve pedagogical efficiency and educational outcomes in complex and diverse learning environments.

Empirical evidence on the impact of GenAI on university learning shows heterogeneous and, in some cases, contradictory results (Richardson et al., 2012). A number of studies have identified positive associations with academic performance, motivation, or learning efficiency. However, other studies have reported null or ambiguous effects, particularly when considering variables such as type of use, academic discipline, or level of digital literacy (Jeynes, 2019). This discrepancy is further exacerbated by the considerable variability in methodological designs, the educational contexts analyzed, the educational variables considered, and the metrics employed to evaluate learning outcomes (Agus, 2022). The predominance of descriptive, perceptual, or exploratory approaches, grounded in surveys and self-reports, engenders limitations in the comparability of findings. Furthermore, previous reviews, predominantly narrative or systematic without quantitative synthesis, lack formal assessments of heterogeneity and bias, which limits an integrated understanding of the phenomenon (Chow & Wehby, 2018).

Recent literature on GenAI in higher education has progressively moved beyond descriptive accounts of adoption toward more analytical discussions focused on its pedagogical, cognitive, and socio-technical implications. Emerging studies frame GenAI not merely as a technological tool, but as a mediator of learning processes that interacts with student agency, instructional design, and institutional governance (Jensen et al., 2025; Qian, 2025; Weng et al., 2024). Within this debate, two dominant perspectives can be identified. On the one hand, techno-optimistic approaches emphasize its potential to enhance feedback, personalization, and learning efficiency. On the other hand, critical perspectives highlight risks related to cognitive offloading, academic integrity, and the reconfiguration of epistemic practices in higher education (Driessens & Pischetola, 2024; Hughes et al., 2025). This tension reflects a broader shift toward understanding GenAI as a socio-technical phenomenon whose educational impact is contingent upon pedagogical integration rather than technological capability alone.

Despite these advances, the existing body of research remains fragmented and largely dominated by qualitative, perceptual, or context-specific empirical studies, which limits the cumulative development of knowledge and the ability to establish generalizable patterns. Furthermore, prior reviews have predominantly focused on mapping trends, perceptions, or adoption factors, without providing robust quantitative synthesis capable of estimating the magnitude and variability of effects across contexts (Lee & Moore, 2024; Weng et al., 2024). In this sense, the present study contributes to advancing the literature by integrating these dispersed findings through a correlational meta-analytic approach, enabling a more rigorous examination of the strength, consistency, and conditional nature of the relationship between GenAI and university educational outcomes. This analytical positioning allows the study to move beyond descriptive accounts and directly engage with ongoing debates regarding the effectiveness, variability, and pedagogical implications of GenAI in higher education.

The extant literature pertaining to the impact of GenAI on university learning reveals significant empirical and methodological gaps, in contrast to other educational fields where meta-analyses have allowed for the estimation of overall effects and the analysis of variability across studies (Korpershoek et al., 2020; Orhan, 2022). To date, there have been no attempts to produce quantitative estimates that integrate effect size in a comparable way across studies. This has prevented determination of the true magnitude of the observed associations. By contrast, educational domains have been extensively synthesized using meta-analytic approaches (Lei et al., 2018; Ma et al., 2016). Furthermore, heterogeneity across studies has not been systematically assessed, despite the high variability in designs, contexts, and metrics. A paucity of attention has been paid to sensitivity and robustness analyses, in addition to an insufficient treatment of publication bias. This has hindered a rigorous quantitative integration of the available empirical evidence.

In this context, the objective of this research is to estimate the overall relationship between GenAI and university educational outcomes through a correlational meta-analysis. In order to achieve this objective, a series of questions have been devised to guide the meta-analysis of the available empirical evidence.

1. To what extent are the results consistent and robust under different analytical decisions and sensitivity procedures?
2. Is there evidence of publication bias that could affect the estimation of the relationship between GenAI and university educational outcomes?
3. What is the magnitude and direction of the overall relationship between GenAI and university educational outcomes estimated through a correlational meta-analysis?
4. Are significant differences observed in the estimated relationship when considering subgroup analyses based on the characteristics of the included studies?

METHODOLOGY

The research employs a quantitative approach through a correlational meta-analysis and adheres to the PRISMA 2020 methodology, the international reporting standard for systematic reviews (Page et al., 2021). This approach is intended to ensure transparency, reproducibility, and methodological rigor throughout the process. PRISMA 2020 is employed exclusively as a reporting and quality assessment framework; it is not utilized as an analytical technique. This enables systematic and traceable documentation of the stages of identification, selection, evaluation, and synthesis of the available empirical evidence.

Inclusion and Exclusion Criteria

The establishment of eligibility criteria was undertaken with the objective of ensuring the inclusion of studies that were both relevant and methodologically consistent, with the ultimate aim of conducting a correlational meta-analysis. Empirical quantitative research conducted in higher education that analyzed the use of GenAI in university contexts and reported quantitative measurements associated with educational outcomes was included, considering only academic publications that explicitly examined this relationship.

The exclusion process was executed in three successive phases. In the initial phase, indexing errors, duplicates, conference proceedings, editorials, book chapters, and documents deemed irrelevant due to their format or thematic focus were eliminated. In the subsequent phase, studies that did not provide full-text access were excluded from the analysis. In the third phase, articles lacking quantitative measurements of the analyzed relationship, which prevented their inclusion in the correlational synthesis, were discarded.

Sources of Information

The present study was conducted using the Scopus and Web of Science databases, which were selected on the basis of their broad multidisciplinary coverage, editorial rigor, and the quality of their indexing processes. This facilitates the systematic retrieval of relevant empirical research in education, technology, and the social sciences with consistent and reliable bibliographic records. The combined use of both databases reflects their complementary nature and helps to reduce the effects of omission bias, thereby increasing the comprehensiveness of the identified evidence, according to comparative analyses of bibliographic coverage (Culbert et al., 2025).

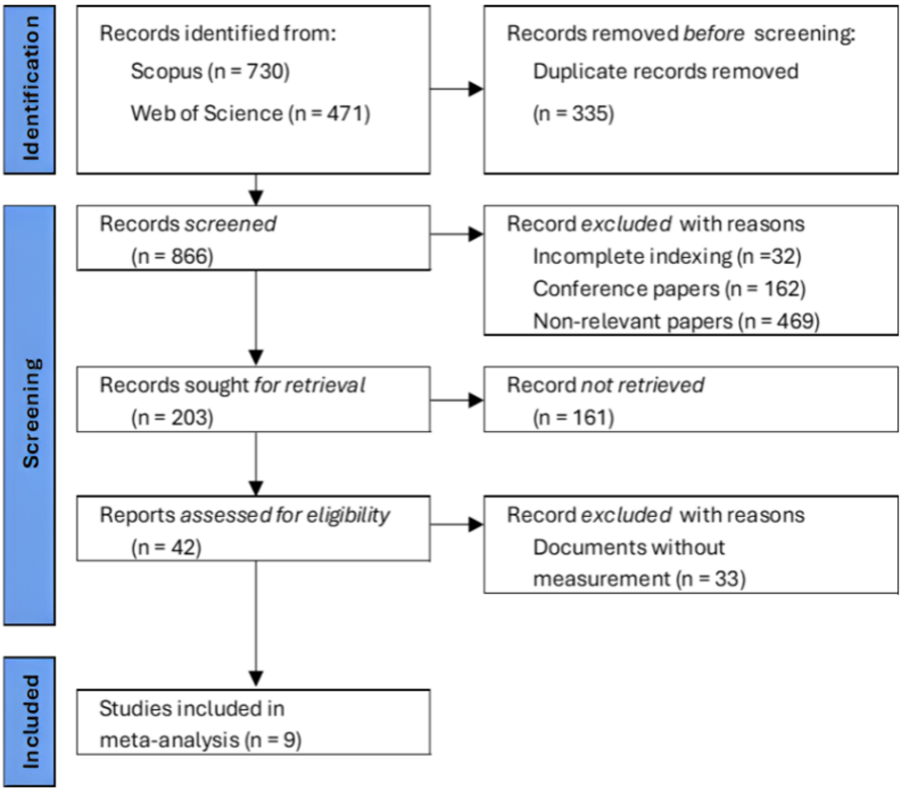

Figure 1. PRISMA flow diagram (prepared by the authors based on Page et al., 2021)

Search Strategy

The search strategy was structured based on the inclusion criteria and defined using a specific equation for each database, with the syntax and search fields being adjusted accordingly. In Scopus, the following search was conducted: TITLE ("generative artificial intelligence" OR "generative AI") AND TITLE ("higher education" OR "university" OR "learning outcomes"). In Web of Science, an equivalent equation was employed, with the relevant fields being adapted: TS = ("generative artificial intelligence" OR "generative AI") AND TS = ("higher education" OR university OR "learning outcomes"), with consideration given to the correspondence with AK = for author keywords. This design enabled the search to concentrate on relevant empirical studies on GenAI and educational outcomes.

Selection Process

The selection process for the study was conducted in accordance with the PRISMA 2020 standard, encompassing the phases of identification, screening, eligibility, and inclusion. **Figure 1** presents a flowchart that summarizes the number of records identified, excluded, and finally included at each stage. The procedure facilitated the systematic documentation of selection decisions, ensuring the traceability, methodological transparency, and reproducibility of the review process.

Data Processing and Statistical Analysis

The data processing was conducted through a systematic procedure of recording, cleansing, and organising the extracted information using Microsoft Excel, followed by statistical analysis of the correlational meta-analysis in RStudio, where the combined effects were estimated and variability between studies was assessed. Furthermore, graphical representations such as forest plots and funnel plots were generated to facilitate the interpretation of the results. When necessary, the statistics reported in the primary studies were converted and standardized to ensure their comparability and integration into the analysis.

Table 1. Evaluation with AMSTAR 2

Authors	Design	Sample	Measure	Overall
Yusuf et al. (2024)	Yes	Yes	Yes	High
Cacho (2024)	Yes	Yes	No	Moderate
Meakin (2024)	No	Yes	No	Low
O'Dea et al. (2026)	Yes	Yes	Yes	High
Ardito (2025)	No	Yes	No	Low
Oc et al. (2025)	Yes	Yes	Yes	High
Haroud and Saqri (2025)	Yes	Yes	Yes	High
Waluyo and Kusumastuti (2024)	Yes	Yes	Yes	High
Perezchica-Vega et al. (2024)	Yes	Yes	No	Moderate
de Fine Licht (2024)	No	Yes	No	Low
Khlaif et al. (2024)	Yes	Yes	Yes	High

Table 2. Characteristics of the included studies

Authors	Dependent variable	Independent variable	n_i	r_i	Variable
Meakin (2024)	GenAI engagement (Thailand EFL students)	English learning outcomes (engagement, confidence, GPA)	28	0.87	Assistive use
Waluyo and Kusumastuti (2024)	GenAI engagement (student acceptance in English learning)	Academic performance (GPA)	28	0.60	Assistive use
Ardito (2025)	GenAI adoption in assessments	Student use behavior	353	0.70	Productive/ creative use
de Fine Licht (2024)	GenAI adoption in assessments	Student use behavior	353	0.55	Productive/ creative use
Khlaif et al. (2024)	Adoption of GenAI tools for student assessment (UTAUT constructs: performance expectancy, effort expectancy, social influence, hedonic motivation)	Behavioral intention and actual use of AI in assessment	358	0.60	Productive/ creative use
Yusuf et al. (2024)	Gender (male vs. female)	Frequency and type of GenAI chatbot use	2,692	0.28	Mixed use
Cacho (2024)	GenAI acceptance (Cengiz & Peker, 2025)	AI anxiety	494	0.22	Mixed use
O'Dea et al. (2026)	GenAI perceptions (Moroccan higher education)	Academic performance & creativity	572	0.52	Mixed use
Perezchica-Vega et al. (2024)	GenAI perceptions (students vs. teachers, Morocco)	Academic performance & creativity	286	0.93	Mixed use

Evaluation of Methodological Quality

The methodological quality of the included studies was assessed using an adaptation of the AMSTAR-2 instrument, considering criteria related to study design, measurement validity, and the adequacy of statistical analyses (Shea et al., 2017). The results of the study are presented in summary form in **Table 1** as an overall assessment without automatic exclusion of studies, consistent with the exploratory nature of the field. Furthermore, a risk of bias assessment was conducted to identify potential limitations associated with sample selection, measurement procedures, and analytical approaches used in the primary studies.

RESULTS

The studies incorporated within the scope of this research, as illustrated in **Table 2**, encompass a diverse array of university contexts, dependent and independent variables, and modes of utilization of GenAI. The analyzed works examine associations between perceptions, acceptance, anxiety, commitment, and adoption of GenAI and various educational outcomes, using quantitative designs with heterogeneous sample sizes and approaches to assistive, mixed, and productive use.

A sensitivity and robustness analysis was performed to assess the stability of the results under different analytical decisions using a random-effects model with a REML estimator (k = 9), in which a significant overall

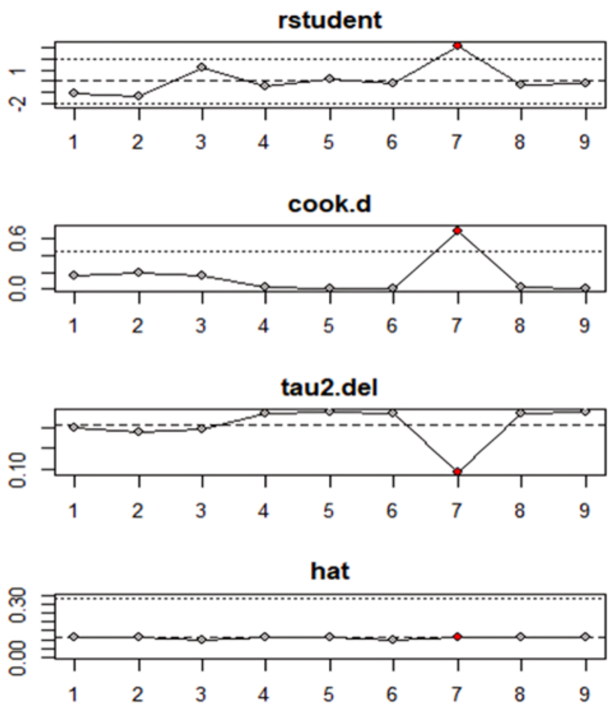

Figure 2. Sensitivity and robustness analysis (the authors' own elaboration)

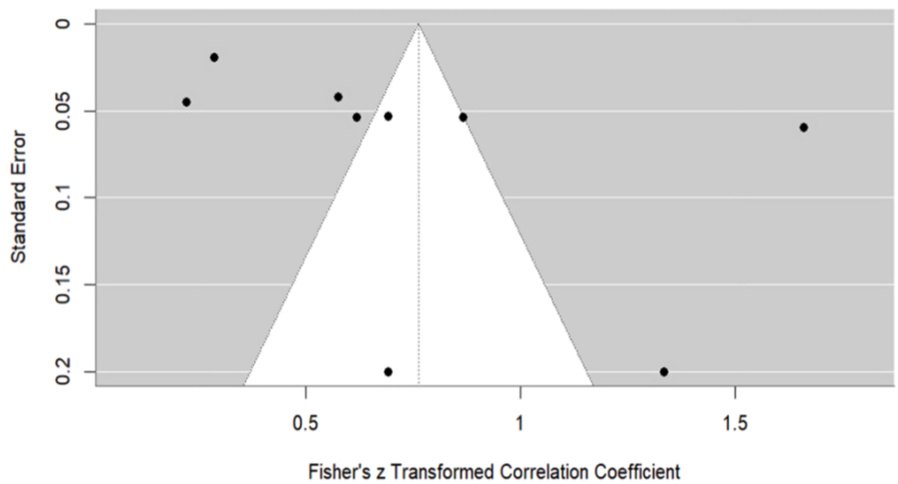

Figure 3. Publication bias (the authors' own elaboration)

effect was observed (r = 0.7627; SE = 0.1554; z = 4.91; $p < .0001$; 95% confidence interval [CI] [0.4581, 1.0673]). The influence analysis identified one study with greater relative weight (case 7), without altering the direction or statistical significance of the estimated effect. **Figure 2** presents the graphical diagnoses associated with the model's stability and influence.

The potential presence of publication bias was assessed using a funnel plot corresponding to the included studies (k = 9), presented in **Figure 3**. The graph displays Fisher's z-transformed correlation coefficients with approximate values ranging from 0.3 to 1.6, along with standard errors ranging from 0.02 to 0.20. Each point on the graph represents an effect estimate distributed according to the precision level of each study.

Furthermore, **Figure 4** presents the main forest plot for the overall assessment of the relationship between GenAI and educational outcomes, integrating k = 9 studies using a random-effects model. The graph displays a combined effect size r = 0.7627 with a 95% CI [0.4581, 1.0673], in conjunction with the individual estimates and their respective CIs. The observed heterogeneity is high (I^2 = 98.92%; τ^2 = 0.2078; τ = 0.4558), reflecting significant variability between studies and supporting the need for further subgroup analysis.

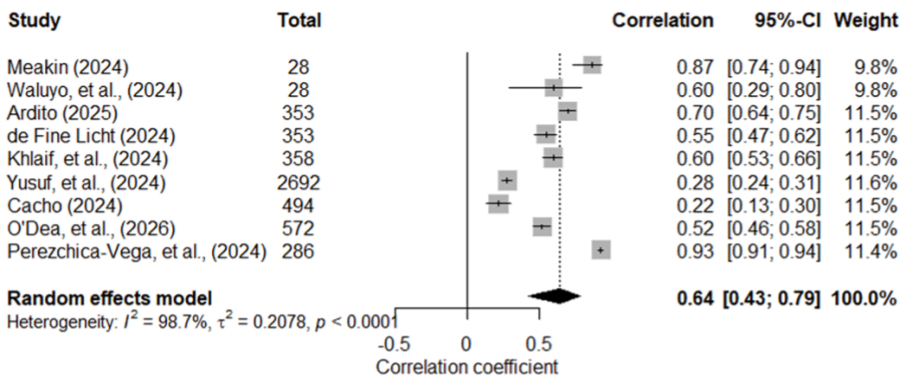

Figure 4. Overall estimate of the relationship (the authors' own elaboration)

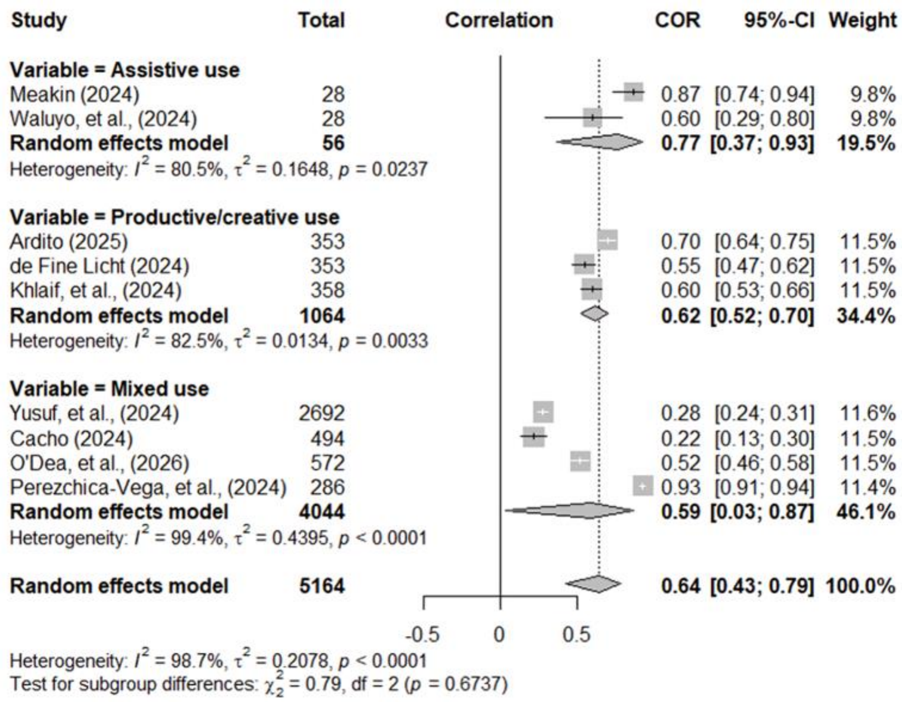

Figure 5. Subgroup and moderator analysis (the authors' own elaboration)

Figure 5 presents the forest plot of the correlational meta-analysis with subgroup evaluation according to the type of use of GenAI: mixed use, assistive use, and productive/creative use. The analysis incorporates a total of nine studies, encompassing a sample size of 5,164 participants. The overall random-effects model reports a pooled correlation of r = 0.64 with a 95% CI [0.43, 0.79], along with high heterogeneity expressed by I^2 = 98.7%, τ^2 = 0.2078, and p < 0.0001.

Moreover, the observed heterogeneity does not appear to be explained by the subgrouping. The statistical analysis indicates that there is a lack of significant variation between the subgroups (χ^2 = 0.79; df = 2; p = 0.6737), suggesting that the differences observed are not statistically significant. This is further supported by the findings, which reveal that there is a lack of statistically relevant variation between mixed use (I^2 = 99.4%), assistive use (I^2 = 80.5%), and productive/creative use (I^2 = 82.5%). Consequently, variability remains high irrespective of the type of use considered.

Figure 6 presents the subgroup analysis based on the type of dependent variable, distinguishing between academic outcomes and artificial intelligence (AI) use and adoption. The subgroup corresponding to academic outcomes includes four studies (n = 914) with a pooled effect size of r = 0.79 [0.50, 0.92] and heterogeneity of I^2 = 98.7%, τ^2 = 0.2647, p < 0.0001. The subgroup corresponding to AI use and adoption includes five studies (n = 4,250) with a pooled effect size of r = 0.49 [0.29, 0.65] and heterogeneity of I^2 = 97.7%, τ^2 = 0.0732, p < 0.0001.

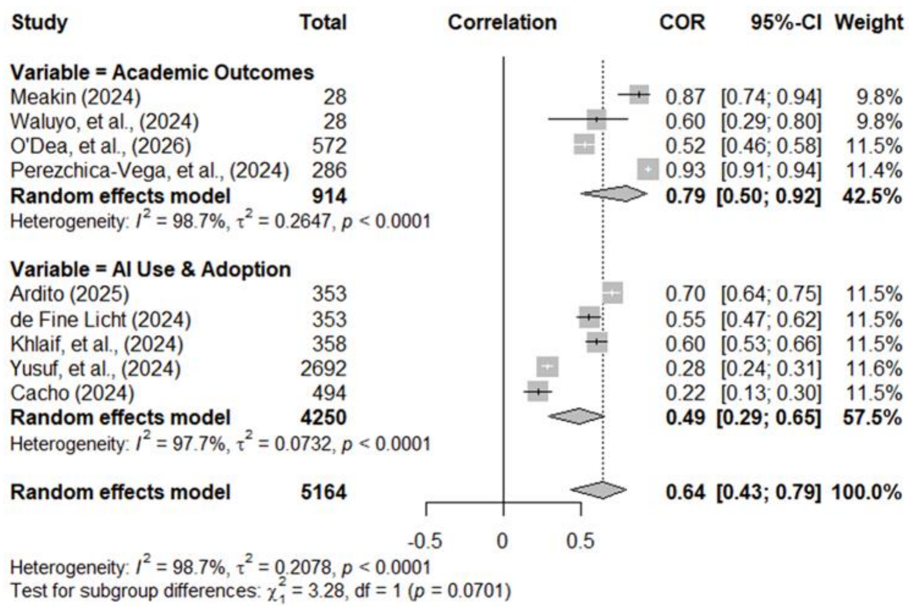

Figure 6. Subgroup and moderator analysis (the authors' own elaboration)

The overall random effects model reports a pooled correlation of r = 0.64 [0.43, 0.79], with heterogeneity of I^2 = 98.7%, τ^2 = 0.2078, $p < 0.0001$. The test for subgroup differences approaches significance but remains non conclusive (χ^2 = 3.28, df = 1, p = 0.0701).

DISCUSSION

The discussion critically interprets the quantitative evidence on the relationship between GenAI and university educational outcomes. The findings of the meta-analysis are examined, compared with previous studies, and a future research agenda is proposed. The theoretical, political, and practical implications are addressed, along with the main methodological and empirical limitations of the study.

Interpretive Synthesis of the Results

The meta-analysis reveals a positive and statistically significant relationship between the utilization of GenAI and university educational outcomes, accompanied by substantial heterogeneity. This finding suggests that GenAI does not function as a uniform or linear educational factor, but rather as a contingent resource whose effect is contingent on the mode of pedagogical integration, the type of academic task, and the institutional frameworks that regulate its implementation. The existence of a high overall effect alongside a wide dispersion of effect sizes indicates that the benefits are neither generalizable nor automatically replicable.

The magnitude of the estimated effect size (r ≈ 0.76) should be interpreted with particular caution, as it is considerably higher than those typically reported in meta-analyses within educational research. While this may suggest a strong association between GenAI and educational outcomes, it is important to consider that such a magnitude may be influenced by the limited number of studies, the high heterogeneity observed, and the inclusion of conceptually diverse variables. In this context, the effect size should not be interpreted as evidence of a uniform or consistently strong impact, but rather as an aggregated indicator reflecting an emerging and methodologically fragmented field. Consequently, the results should be understood as indicative of a potentially significant relationship that requires further validation through more homogeneous, large-scale, and methodologically comparable studies.

The persistence of heterogeneity after the subgroup analysis demonstrates that the mode of use only partially explains the observed variability. This suggests the intervention of pedagogical, psychological, ethical, and institutional mechanisms that are not always empirically captured. This interpretation is consistent with the extant literature, which conceptualizes GenAI as a socio-technical technology and not as an isolated instructional tool.

The subgroup analyses presented in **Figure 5** and **Figure 6** provide two complementary perspectives on the variability of the estimated effects, each capturing a different dimension of the phenomenon. **Figure 5** organizes the analysis according to the type of GenAI use, distinguishing between assistive use, productive or creative use, and mixed use. The results indicate a clear gradient, where assistive use shows the highest effect size ($r = 0.77$), followed by productive or creative use ($r = 0.62$) and mixed use ($r = 0.59$). This pattern suggests that more structured and task oriented applications of AI tend to produce stronger associations with educational outcomes than broader or less defined usage contexts.

In contrast, **Figure 6** classifies the studies according to the type of dependent variable, differentiating between academic outcomes and AI use and adoption. This configuration reveals a consistent pattern in which academic outcomes exhibit higher effect sizes ($r = 0.79$) compared to AI use and adoption ($r = 0.49$), indicating that GenAI has a stronger relationship with direct learning performance indicators than with behavioral or adoption related constructs. The convergence of these patterns across both **Figure 5** and **Figure 6** suggests that both the nature of AI use and the way outcomes are operationalized act as relevant sources of variation in the estimated effects.

However, in both subgroup structures, the absence of statistically significant differences and the persistence of high heterogeneity levels indicate that these classifications only partially explain the observed variability. In **Figure 5**, the subgroup differences are clearly non- significant ($p = 0.6737$), while in **Figure 6** they approach significance but remain non conclusive ($p = 0.0701$). This suggests that the variability in effect sizes is driven by additional unobserved factors, such as differences in instructional design, intensity of AI integration, or contextual characteristics of the learning environment. Therefore, the relationship between GenAI and educational outcomes should be interpreted as inherently context dependent and shaped by multiple interacting conditions rather than by isolated subgroup distinctions.

From this standpoint, the most significant impacts related to assistive use are associated with structured digital tutoring models, in which AI plays a cognitive and metacognitive mediation role without substituting student agency, as evidenced by experiences with pedagogically validated digital tutors (Reicher et al., 2025). In contrast, the more moderate and heterogeneous effects of mixed use reflect less regulated contexts, where the combination of support, automation, and autonomous production introduces tensions between learning, technological dependence, and academic integrity. This problem has been widely documented in studies on AI assessment and instructional design (Ilieva et al., 2025). Conversely, technology acceptance models furnish supplementary explanatory elements. The findings of the meta-analysis are consistent with the extant literature identifying intention to use, technological self-efficacy, and perceived benefits as central determinants of the educational impact of GIA on students and teachers (Alotaibi, 2026; Bamasoud et al., 2025; Tbaishat et al., 2025).

The considerable heterogeneity observed among the studies can be interpreted as reflecting contextual variations in the extent of institutional support, the quality of teacher training, and the presence of clear regulatory frameworks. These factors are recognized as being critical for effective adoption, yet they are seldom operationalized in a consistent manner in correlational studies. From a broader standpoint, the findings are consistent with critical approaches that emphasize the inextricable linkage of the educational effects of GIA with ethical, regulatory and cultural considerations. The documented heterogeneity of institutional responses to GenAI, as evidenced by analyses of university policies and governance frameworks, helps to elucidate why seemingly analogous contexts can yield disparate educational outcomes (Dabis & Csáki, 2024; Smith et al., 2026; Tong et al., 2025). In this sense, the high heterogeneity observed is not a weakness of the meta-analysis, but rather an empirical indicator that AI-mediated educational transformation is a structurally unequal process, conditioned by pedagogical, regulatory, and cultural decisions more than by the technology itself (Fern, 2024).

Comparison with Other Studies

The results of the meta-analysis demonstrate partial convergence with the evidence derived from previous meta-analyses concerning psychoeducational and technological factors associated with learning. As evidenced by numerous studies examining school belonging, student engagement, and critical thinking, a positive and statistically significant relationship with academic outcomes has been identified, albeit with heterogeneous magnitudes (Korpershoek et al., 2020; Lei et al., 2018; Orhan, 2022). However, the effect size estimated in this study (r = 0.76) consistently exceeds the values reported in these meta-analyses, suggesting a potential amplification of GenAI on previously documented cognitive, motivational, and self-regulatory processes. In contrast to meta-analyses that concentrate on relatively stable variables, such as parental involvement or engagement, the high heterogeneity observed indicates that the effect of GenAI is markedly contingent on the context and the mode of use. The study's distinctive contribution lies in providing a specific quantitative synthesis on GenAI, incorporating robustness, publication bias, and subgroup analyses—elements generally absent in narrative reviews or individual studies.

Research Agenda

The synthesized evidence indicates the necessity for a future research agenda that progresses beyond the limits of descriptive and techno-deterministic approaches, and advances towards explanatory models capable of capturing the socio-technical complexity of GenAI in higher education. In light of the identified empirical gaps, it is imperative to develop more sophisticated analytical subgroups that transcend broad classifications of use and integrate specific combinations of task type, academic discipline, level of student autonomy, and degree of teacher mediation. This is because the mode of use only partially explained the observed heterogeneity. This disaggregation would facilitate the identification of specific pedagogical configurations associated with differential effects and reduce unexplained residual variability. Furthermore, research should systematically incorporate pedagogical, cognitive, and contextual variables that are scarcely operationalized in the primary literature. Such variables include self-regulated learning, cognitive load, assessment design, curriculum alignment, and institutional ethical frameworks.

The absence of these dimensions has a significant impact on the interpretation of effect sizes, thereby reinforcing the necessity to transition from exclusively correlational designs towards longitudinal and quasi-experimental studies. These novel approaches will facilitate the examination of learning trajectories and the cumulative effects of technology-mediated learning (TML). Furthermore, the adoption of multilevel models that integrate student, course and institutional factors is recommended, in recognition of the fact that the impact of TML emerges from the interaction between these levels. Finally, it is necessary to conceptually differentiate between instrumental, pedagogical, and epistemological uses of TML and to adopt mixed-methods approaches that articulate quantitative evidence with in-depth qualitative analyses in order to clarify underlying mechanisms, ethical tensions, and actual implementation practices currently obscured in available empirical studies.

IMPLICATIONS

The theoretical implications of this meta-analysis demand a rethinking of the dominant models of TML in higher education. The magnitude of the overall effect, coupled with the presence of extremely high heterogeneity, poses a significant challenge to linear approaches that presuppose a direct and homogeneous relationship between the use of GenAI and improved educational outcomes. The extant evidence supports a non-deterministic conception of GenAI as a sociotechnical mediator, the impact of which is shaped by the interaction between pedagogical design, student agency, and institutional context. This necessitates a shift in theoretical frameworks from tool-centered approaches to ones that integrate cognitive, metacognitive, motivational, and normative processes. These frameworks must also recognize that technology amplifies pre-existing educational dynamics rather than producing autonomous effects.

In terms of policy, the results indicate the necessity for institutional and regulatory policies that are grounded in empirical evidence, as opposed to reactive or prohibitionist responses. The substantial heterogeneity observed across studies suggests that generic regulations on GIA may be inadequate and potentially counterproductive. Consequently, there is a necessity for ethical and pedagogical frameworks guided by explicit educational objectives. It is imperative that these frameworks incorporate criteria of transparency, accountability, and equity, and promote the strengthening of institutional capacities in teacher training, assessment redesign, and data and algorithm governance. The absence of such frameworks has been demonstrated to contribute to the observed dispersion of effects and limits the transformative potential of GIA.

From a pragmatic standpoint, the findings suggest that educators, instructional designers and academic administrators must transcend mere instrumental approaches and adopt a strategic and contextualized utilization of GIA. The extant evidence suggests that there should be a prioritization of pedagogical integrations where AI acts as cognitive and metacognitive support, aligned with authentic tasks and coherent evaluation criteria. Such integrations should avoid indiscriminate or substitute uses, and the institutional decision-making process should be based on contextual diagnoses that define when, how and for what purpose AI provides educational value. It is important to note that the impact of AI depends less on the technology itself than on the pedagogical and organizational conditions that structure it.

Limitations

The study is subject to certain limitations that must be considered when interpreting the results, since the synthesized evidence is exclusively derived from correlational studies, which prevents the establishment of causal relationships between the utilization of GenAI and educational outcomes. Moreover, the heterogeneity among studies is so great that the generalizability of the estimated effects is restricted. Another important aspect to consider is the extremely high heterogeneity observed across the included studies ($I^2 \approx 99\%$). This heterogeneity does not constitute a limitation of the meta-analytic procedure itself but rather reflects the intrinsic variability of the emerging field of GenAI in higher education, where differences in study design, variables, contexts, and measurement approaches are substantial. The inclusion of conceptually diverse constructs, such as academic performance, perceptions, adoption, and affective variables, responds to the current state of the literature, in which educational outcomes are operationalized in heterogeneous ways. However, this variability must be carefully considered when interpreting the aggregated effect size, as it limits the comparability of findings and suggests that the estimated relationship should be understood as a general tendency rather than a uniform or directly comparable effect.

The availability and quality of primary reports influenced the process of information extraction and constrained the systematic incorporation of relevant moderators. Finally, the restriction to specific databases may have resulted in the exclusion of relevant non-indexed literature, thereby reducing the overall coverage of the analyzed field.

An additional limitation concerns the relatively small number of studies included in the meta-analysis ($k = 9$). This was the result of a deliberate and restrictive selection process that prioritized methodological rigor and statistical suitability over the inclusion of a larger number of studies. Only empirical investigations reporting quantitative associations suitable for correlational synthesis were retained, excluding a substantial body of descriptive, qualitative, or non-comparable research. While this approach enhances the internal validity and comparability of the estimated effects, it inevitably reduces the breadth of the evidence base. Consequently, the findings should be interpreted with caution, as the limited number of studies may affect the stability of the estimates and the generalizability of the results. Future research would benefit from a broader accumulation of statistically compatible studies to strengthen the robustness of meta-analytic conclusions.

CONCLUSIONS

The findings confirm that GenAI constitutes a relevant and influential factor in university learning, although its effects cannot be interpreted as uniform or deterministic. The meta analytic evidence demonstrates a consistent positive association between GenAI and educational outcomes, while also revealing substantial variability across studies. The subgroup analyses presented in **Figure 5** and F**igure 6** provide complementary evidence showing that both the type of AI use and the type of outcome measured influence the magnitude of the observed effects. Specifically, stronger associations are observed in more structured forms of use such as assistive applications and in direct academic outcomes, while lower effect sizes are found in broader usage contexts and behavioral or adoption related variables. However, these differences are not statistically significant and are accompanied by high levels of heterogeneity, indicating that they do not fully explain the variability of the results.

In this sense, the study demonstrates that the impact of GenAI depends less on the technology itself and more on how it is used, implemented, and integrated within specific educational contexts. The meta-analysis provides a rigorous overall estimate while highlighting the complex and context-dependent nature of technological effects in education.

Author contributions: AV-A: conceptualization, formal analysis, methodology, supervision, writing–original draft, and writing–review & editing; **EOSM:** data curation, formal analysis, methodology, and writing–review & editing; **LGPS:** data curation, validation, and writing–review & editing; **GAML, SAG, & WL-C:** writing–review & editing; & **PAR-C:** data curation and writing–review & editing. All authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Ethics declaration: This study is a correlational meta-analysis based exclusively on previously published studies; therefore, no primary data were collected directly from patients, students, teachers, or other human participants. Since the analysis used only aggregate data reported in public academic publications, informed consent was not applicable. No personal, sensitive, identifiable, or confidential information was accessed, stored, processed, or disclosed at any stage of the study.

AI statement: Generative AI tools were used only to support language editing, grammar correction, and improvement of clarity and readability of the manuscript. These tools were not used to generate the research idea, design the methodology, collect or analyze data, interpret results, create conclusions, or make substantive academic decisions. All content was carefully reviewed, verified, and approved by the authors, who take full responsibility for the final version of the manuscript.

Declaration of interest: The authors declared no competing interests.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Agus, A. R. R. (2022). Meta-analysis: Correlation study between science process skills and learning outcomes. *Journal of Digital Learning and Education, 2*(3), 196-202. https://doi.org/10.52562/jdle.v2i3.427

Alotaibi, N. (2026). Faculty acceptance of generative AI in higher education: A meta-analysis of TAM and UTAUT studies (2021-2025). *International Journal of Higher Education, 15*(1), 1-14. https://doi.org/10.5430/ijhe.v15n1p1

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

Bamasoud, D. M., Mohammad, R., & Bilal, S. (2025). Adopting generative AI in higher education: A dual-perspective study of students and lecturers in Saudi universities. *Big Data and Cognitive Computing, 9*(10), Article 264. https://doi.org/10.3390/bdcc9100264

Beauvais, A. M., Stewart, J. G., DeNisco, S., & Beauvais, J. E. (2014). Factors related to academic success among nursing students: A descriptive correlational research study. *Nurse Education Today, 34*(6), 918-923. https://doi.org/10.1016/j.nedt.2013.12.005

Cacho, R. M. (2024). Integrating generative AI in university teaching and learning: A model for balanced guidelines. *Online Learning, 28*(3), 55-81. https://doi.org/10.24059/olj.v28i3.4508

Cengiz, S., & Peker, A. (2025). Generative artificial intelligence acceptance and artificial intelligence anxiety among university students: The sequential mediating role of attitudes toward artificial intelligence and literacy. *Current Psychology, 44*(9), 7991-8000. https://doi.org/10.1007/s12144-025-07433-7

Chow, J. C., & Wehby, J. H. (2018). Associations between language and problem behavior: A systematic review and correlational meta-analysis. *Educational Psychology Review, 30*(1), 61-82. https://doi.org/10.1007/s10648-016-9385-z

Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics, 130*(4), 2475-2492. https://doi.org/10.1007/s11192-025-05293-3

Dabis, A., & Csáki, C. (2024). AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanities and Social Sciences Communications, 11*, Article 1006. https://doi.org/10.1057/s41599-024-03526-z

de Fine Licht, K. (2024). Generative artificial intelligence in higher education: Why the 'banning approach' to student use is sometimes morally justified. *Philosophy & Technology, 37*, Article 113. https://doi.org/10.1007/s13347-024-00799-9

Driessens, O., & Pischetola, M. (2024). Danish university policies on generative AI: Problems, assumptions and sustainability blind spots. *MedieKultur: Journal of Media and Communication Research, 40*(76), 31-52. https://doi.org/10.7146/mk.v40i76.143595

Ermold, C. M. (2011). *A correlational study of student achievement and school climate* [PhD thesis, Walden University]. https://www.proquest.com/openview/7b43afb7abee4c171aa0ae0e2a373355/1?pq-origsite=gscholar&cbl=18750

Fern, J. (2024). A more-than-human ecology: Evolving generative artificial intelligence in higher education. *Education Sciences, 14*(10), Article 1102. https://doi.org/10.3390/educsci14101102

Haider, S. Z., & Hussain, A. (2014). Relationship between teacher factors and student achievement: A correlational study of secondary schools. *US-China Education Review A, 4*(7), 465-480. https://www.davidpublisher.com/Public/uploads/Contribute/551233e421b66.pdf

Haroud, S., & Saqri, N. (2025). Generative AI in higher education: Teachers' and students' perspectives on support, replacement, and digital literacy. *Education Sciences, 15*(4), Article 396. https://doi.org/10.3390/educsci15040396

Hughes, L., Malik, T., Dettmer, S., Al-Busaidi, A. S., & Dwivedi, Y. K. (2025). Reimagining higher education: Navigating the challenges of generative AI adoption. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-025-10582-6

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

Iqbal, S., Akram, M., & Mushtaq, I. (2021). Relationship between stress and educational performance of university students: A correlational research study. *Review of Education, Administration & Law, 4*(4), 805-811. https://doi.org/10.47067/real.v4i4.198

Ishak, Okilanda, A., Permadi, A. A., Tjahyanto, T., Prabowo, T. A., Rozi, M. F., Suganda, M. A., & Suryadi, D. (2023). Correlational study: Sports Students' special test results and basic athletic training learning outcomes. *Retos, 49*, 519–524. https://doi.org/10.47197/retos.v49.98820

Jensen, L. X., Buhl, A., Sharma, A., & Bearman, M. (2025). Generative AI and higher education: A review of claims from the first months of ChatGPT. *Higher Education, 89*(4), 1145-1161. https://doi.org/10.1007/s10734-024-01265-3

Jeynes, W. H. (2019). A meta-analysis on the relationship between character education and student achievement and behavioral outcomes. *Education and Urban Society, 51*(1), 33-71. https://doi.org/10.1177/0013124517747681

Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences, 14*(10), Article 1090. https://doi.org/10.3390/educsci14101090

Korpershoek, H., Canrinus, E. T., Fokkens-Bruinsma, M., & De Boer, H. (2020). The relationships between school belonging and students' motivational, social-emotional, behavioural, and academic outcomes in secondary education: A meta-analytic review. *Research Papers in Education, 35*(6), 641-680. https://doi.org/10.1080/02671522.2019.1615116

Lee, S. S., & Moore, R. L. (2024). Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learning, 28*(3), 82-106. https://doi.org/10.24059/olj.v28i3.4593

Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal, 46*(3), 517-528. https://doi.org/10.2224/sbp.7054

Ma, X., Shen, J., Krenn, H. Y., Hu, S., & Yuan, J. (2016). A meta-analysis of the relationship between learning outcomes and parental involvement during early childhood education and early elementary education. *Educational Psychology Review, 28*(4), 771-801. https://doi.org/10.1007/s10648-015-9351-1

Meakin, L. (2024). Exploring the impact of generative artificial intelligence on higher education students' utilization of library resources: A critical examination. *Information Technology and Libraries, 43*(3), 1-13. https://doi.org/10.5860/ital.v43i3.17246

O'Dea, X., Tsz Kit Ng, D., O'Dea, M., & Shkuratskyy, V. (2026). Factors affecting university students' generative AI literacy: Evidence and evaluation in the UK and Hong Kong contexts. *Policy Futures in Education, 24*(1), 13-34. https://doi.org/10.1177/14782103241287401

Oc, Y., Gonsalves, C., & Quamina, L. T. (2025). Generative AI in higher education assessments: Examining risk and tech-savviness on student's adoption. *Journal of Marketing Education, 47*(2), 138-155. https://doi.org/10.1177/02734753241302459

Orhan, A. (2022). The relationship between critical thinking and academic achievement: A meta-analysis study. *Psycho-Educational Research Reviews, 11*(1), 283-299. https://doi.org/10.52963/PERR_Biruni_V11.N1.18

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Perezchica-Vega, J. E., Sepúlveda-Rodríguez, J. A., & Román-Méndez, A. D. (2024). Inteligencia artificial generativa en la educación superior: Usos y opiniones de los profesores [Generative artificial intelligence in higher education: Uses and opinions of professors]. *European Public & Social Innovation Review, 9*, 1-20. https://doi.org/10.31637/epsir-2024-593

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends, 69*, 1105-1120. https://doi.org/10.1007/s11528-025-01100-1

Reicher, H., Frenkel, Y., Lavi, M. J., Nasser, R., Ran-milo, Y., Sheinin, R., Shtaif, M., & Milo, T. (2025). A generative AI-empowered digital tutor for higher education courses. *Information, 16*(4), Article 264. https://doi.org/10.3390/info16040264

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353-387. https://doi.org/10.1037/a0026838

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ, 358*, Article j4008. https://doi.org/10.1136/bmj.j4008

Smith, S. M., Tate, M., Freeman, K., Walsh, A., Ballsun-Stanton, B., & Lane, M. (2026). A university framework for the responsible use of generative AI in research. *Journal of Higher Education Policy and Management, 48*(1), 17-36. https://doi.org/10.1080/1360080X.2025.2509187

Tbaishat, D., Amoudi, G., & Elfadel, M. (2025). Adapting teaching and learning with existing generative AI by higher education students: Comparative study of Zayed University and King Abdulaziz University. *Computers and Education: Artificial Intelligence, 8*, Article 100421. https://doi.org/10.1016/j.caeai.2025.100421

Tong, A., Zainol, Z., Chong, T. S., & Renganathan, K. (2025). AI governance on young consumers in higher education: A content analysis of policies for generative AI. *Young Consumers: Insight and Ideas for Responsible Marketers, 26*(5), 865-885. https://doi.org/10.1108/YC-10-2024-2303

Waluyo, B., & Kusumastuti, S. (2024). Generative AI in student English learning in Thai higher education: More engagement, better outcomes? *Social Sciences & Humanities Open, 10*, Article 101146. https://doi.org/10.1016/j.ssaho.2024.101146

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

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*, Article 21. https://doi.org/10.1186/s41239-024-00453-6