

The Impact of Artificial Intelligence Applications on University Teaching Quality and Students' Self-Directed Learning Skills

Mutaib Al-Otaibi , Department of Basic Sciences and Humanities, Faculty of Arts and Sciences, Amman Arab University, Amman, Jordan, Oricd No: <https://orcid.org/0000-0003-4306-4517>, Email: dr.mutaib68@aau.edu.jo.

Abstract

Artificial intelligence (AI) applications are rapidly reshaping university teaching through adaptive tutoring, generative dialogue, automated feedback, learning analytics, predictive support, and AI-assisted assessment. Yet the educational value of these applications depends not merely on access to technology but on how they are embedded in pedagogy. This study examines the impact of AI applications on **university teaching quality** and students' **self-directed and self-regulated learning skills**. It uses a systematic integrative review and a secondary quantitative synthesis of verified, peer-reviewed evidence. The principal quantitative corpus comprises 85 higher-education studies, 387 dependent effect-size estimates, and 9,564 participants published between 2015 and 2025. Descriptive analyses were reconstructed from publicly available supplementary data, while inferential conclusions were based on published three-level and random-effects meta-analytic estimates; no synthetic data were used. AI feedback had a medium positive average effect on self-regulated learning (Hedges' $g = 0.509$), but its 95% prediction interval ranged from -0.661 to 1.678 , indicating that effects are highly conditional. Generative-AI interventions also improved academic performance ($g = 0.683$, 95% CI $[0.17, 1.19]$), while ChatGPT-supported learning improved behavioral, cognitive, and emotional engagement. However, heterogeneity was substantial, AI feedback did not clearly outperform

human feedback, and experimental evidence identified risks of dependence, shallow processing, and “metacognitive laziness.” Teacher support emerged as a decisive pedagogical condition in the academic-performance literature: teacher-supported student–AI interaction produced a large effect ($g = 1.426$), whereas unsupported use produced a near-zero effect ($g = 0.077$). The study concludes that AI can improve university teaching quality and learner autonomy when it functions as a **scaffold within teacher-designed learning**, not as a substitute for disciplinary teaching, human feedback, reflective judgment, or independent performance.

Keywords: artificial intelligence in higher education; generative AI; teaching quality; self-directed learning; self-regulated learning; university students; learning analytics; intelligent tutoring systems; feedback; meta-analysis

1. Introduction

Artificial intelligence has moved from a specialized educational technology to a widely accessible component of university learning. Contemporary applications include intelligent tutoring systems, generative-AI dialogue systems, adaptive courseware, automated writing evaluation, personalized feedback, learning-analytics dashboards, predictive models, and virtual teaching assistants. In the higher-education field, the most prominent application topics have been assessment and evaluation, profiling and prediction, adaptive systems and personalization, and intelligent tutoring systems; but historically, the higher-education field has focused more on technical systems and student outcomes than on educators and pedagogy (Crompton & Burke, 2023; Zawacki-Richter et al., 2019).

The expansion of generative AI has intensified longstanding questions about instructional quality. AI can provide rapid feedback, individualized explanations, repeated low-risk practice, and support outside scheduled class time. At the same time, unreliable content, opaque

reasoning, privacy concerns, academic misconduct, cognitive offloading, and dependency can undermine the learning processes universities are expected to cultivate. The relevant question is therefore not whether AI is inherently beneficial or harmful. It is whether specific AI functions, embedded in specific instructional designs, improve meaningful learning outcomes while preserving students' agency, evaluative judgment, and capacity to learn independently.

Teaching quality is a multidimensional construct. In this study, it refers to the degree to which instructional arrangements enable clear guidance, timely and usable feedback, cognitive engagement, equitable access to support, efficient formative assessment, academic achievement, higher-order thinking, and sustained learner participation. This definition treats student outcomes as important indicators of instructional effectiveness but does not reduce teaching quality to grades alone. The broad tertiary review revealed that 97.0 per cent of the AI in higher education evidence syntheses are focused on teaching and learning support, with only five of the 66 reviews specifically reporting on the use of AI to evaluate TE, highlighting the need for differentiation between the quality of AI supported learning and the use of AI to assess teachers (Bond et al., 2024).

Self-directed learning (SDL) and self-regulated learning (SRL) overlap but are not identical. SDL concerns a learner's broader initiative in diagnosing needs, setting goals, selecting resources, implementing strategies, and evaluating outcomes. SRL concerns the cyclical regulation of cognition, metacognition, motivation, emotion, and behavior during learning tasks. In higher education, both constructs describe the movement from externally managed study toward purposeful learner agency. Panadero's (2017) synthesis of SRL models and research on online university learning show that goal setting, strategic action, monitoring, help seeking, time

and effort management, and reflection are central to successful autonomous learning (Broadbent & Poon, 2015; Panadero, 2017).

The distinction is important because a tool may improve the immediate product without strengthening independent learning. For example, a student may submit a better essay after accepting AI-generated revisions but demonstrate little improvement in knowledge transfer, monitoring accuracy, or unaided writing. Conversely, well-designed prompts, visualizations, and feedback can make learning processes visible and help students become more strategic. The educational effect of AI must therefore be evaluated at both the **performance level** and the **agency level**.

1.1 Problem statement

The empirical literature reports positive, null, and negative effects of AI in university teaching. Reviews show benefits for personalized support, academic achievement, feedback, and engagement, but they also identify theoretical fragmentation, inconsistent measurement, short interventions, weak reporting, and insufficient attention to educators' roles (Banihashem et al., 2025; Bond et al., 2024; Lan & Zhou, 2025). A central research problem is therefore to determine the average direction and magnitude of AI's effects while identifying the pedagogical conditions under which positive outcomes are more or less likely.

1.2 Purpose and research questions

The purpose of this study is to synthesize credible quantitative and qualitative evidence on the effects of AI applications in higher education. It addresses three questions.

Research question	Focus
RQ1	To what extent do AI applications improve university teaching quality, as reflected in academic

	performance, feedback, engagement, and instructional support?
RQ2	To what extent do AI applications develop students' self-directed and self-regulated learning skills?
RQ3	Which technological and pedagogical conditions explain variation in these effects?

1.3 Significance of the study

This study contributes by integrating three evidence streams that are often treated separately: research on teaching effectiveness and academic performance, research on engagement, and research on SDL/SRL. It also uses openly available coded data to describe the composition of a recent 85-study higher-education meta-analytic corpus. The result is a practical evidence model for universities: AI is most educationally valuable when its immediacy and personalization are combined with instructional alignment, feedback literacy, verification, reflective activity, and progressively independent performance.

2. Conceptual and Theoretical Framework

2.1 Self-regulated learning as a cyclical process

SRL is commonly represented as a cycle of forethought, performance, and self-reflection. Prior to the execution of the task, learners examine the task, make decision about goals, elicit prior knowledge, and make efficacy and value judgments. In performance they choose strategies, they track for understanding, they manage their efforts, seek help and control their attention. They also assess the consequences in reflection, assign reasons, adjust strategies, and set up new objectives (Panadero, 2017). Based on the research findings from the various studies in the meta-analysis, it is concluded that SRL explicit training has a positive impact on the SRL strategy use

and motivation of university students, thus supporting the view that autonomous learning is not an innate personal ability but it is teachable (Theobald, 2021).

There are ways in which AI can help at every step. Forethought can assist in breaking down criteria, assessing prerequisite knowledge, proposing plans, and creating practicing routes. It is capable of giving explanations, hints, adaptive problems, feedback, progress visualizations and help seeking on performance. It can offer a reflection that allows work to be compared to criteria, provides questions, summarizes trace data, and reminds learners of their errors. Several studies in educational Chabot's and students' preference for L2 design have highlighted that there must be a balance between L2 learner's activeness, identity and control over L2 designing and decision making, otherwise, it is not considered effective SRL support (Jin et al., 2023; Guan et al., 2025). The qualitative review of 14 higher-education studies indicated potential support on all three phases, but also revealed that the agency could be differently distributed between SRL and human-centered and AI-centered agency, which would reorganize the SRL cycle (Lan & Zhou, 2025).

2.2 Self-determination and learner agency

Self-determination theory suggests that intrinsic-motivation is enhanced by autonomy, competence and relatedness in learning (Deci & Ryan, 2000; Ryan & Deci, 2000). AI can help to increase perceived competence by providing rapid feedback and scaffolding over time, as well as by offering varying feedback on time, pace, examples, and pathways through content; AI can increase autonomy by providing choices over time and pace, examples and pathways through content, and scaffolding when needed. However, apparent autonomy can become dependency if the system performs the learner's core cognitive work. Similarly, automated assistance may

reduce the relational dimension of learning unless it is integrated into dialogue with instructors and peers.

The distinction between **autonomy support** and **task outsourcing** is therefore central. Autonomy support increases the learner's capacity to act. Task outsourcing merely transfers the activity to the system. A useful instructional test is whether students can subsequently explain, critique, adapt, and reproduce the performance without AI support. If they cannot, the tool has improved production efficiency but may not have improved learning autonomy.

2.3 Feedback theory and feedback literacy

Feedback supports learning when it reduces the gap between current and intended performance and when learners can interpret and use it. Automated feedback can be timely, scalable, and consistent, but speed alone does not guarantee educational value. The reviews of automatically generated feedback have found that there is a high degree of variation in the personalization, degree of explanation, learner control, and connection to pedagogical objectives (Deeva et al., 2021). AI feedback must therefore be judged by whether it directs attention to relevant criteria, stimulates diagnosis and revision, and leads to improved future performance.

Feedback literacy is especially important in generative-AI contexts because fluent language can mask factual errors, weak evidence, or superficial reasoning. Students need to compare feedback with task criteria, disciplinary sources, and their own intentions. Teachers must also decide which aspects of feedback can be automated and which require professional judgment, emotional sensitivity, or knowledge of the learner's development.

2.4 An integrated mechanism model

The theoretical framework used in this study proposes that AI affects teaching quality and self-directed learning through five connected mechanisms.

Mechanism	Potential educational contribution	Principal risk if poorly designed
Personalization	Modifies the content, examples, difficulty, pace and feedback to suit the learner's needs.	Algorithmic narrowing or incorrect learner models, or recommendations that are not equal
Immediacy	Minimizes the delay in feedback and provides opportunities for multiple revision and practice.	Fast answers can short-circuit good amount of thought and wrestling.
Visibility	Makes patterns of participation, progress, errors and patterns of strategy use visible.	Surveillance, loss of privacy, reductive metrics or misclassification.
Dialogue	Allows for iterative questioning, explanation, rehearsal and feedback.	Persistent misinformation, belief, leading and addiction.
Scaffolding	Supplies suggestions, cues and/or scaffolding that can be phased out as competence increases.	Independent performance and metacognitive development may not be possible if using permanent support.

This model predicts that AI will have stronger effects when it helps learners perform actions they are ready to internalize and when support is gradually reduced. It predicts weaker or negative effects when AI removes essential reasoning, evaluation, or retrieval practice from the task.

3. Method

3.1 Research design

The study used a **systematic integrative review and secondary quantitative evidence synthesis**. It is not presented as a new de novo meta-analysis. The principal statistical dataset was reconstructed from the publicly available supplementary materials of a three-level meta-

analysis of AI feedback and SRL in higher education (Huang et al., 2026). The corpus contains 85 studies, 387 statistically dependent effect-size estimates, and 9,564 participants.

Complementary meta-analyses, systematic reviews, and primary higher-education studies were used to evaluate academic performance, engagement, instructional quality, mechanisms, and risks.

3.2 Search and evidence identification

The principal corpus searched EBSCOhost education and psychology databases, Web of Science, Scopus, and ProQuest for English-language studies published from January 2015 to December 12, 2025. Its search combined terms for AI technologies, feedback, SRL processes, and higher education and initially identified 6,933 records. The present study additionally searched for recent meta-analyses, systematic reviews, and primary studies using combinations of “artificial intelligence,” “generative AI,” “higher education,” “teaching quality,” “academic performance,” “student engagement,” “self-directed learning,” “self-regulated learning,” “feedback,” and “DOI.” Reporting principles were informed by PRISMA 2020, while recognizing that the present work integrates an existing meta-analytic dataset rather than claiming an independent replication of its complete screening process (Page et al., 2021).

3.3 Eligibility and verification

Evidence was prioritized when it met four conditions: it addressed university or postsecondary education; examined an educational AI application or AI-supported instructional process; reported teaching, learning, engagement, SDL, or SRL outcomes; and possessed verifiable publisher metadata and a DOI. Experimental and quasi-experimental studies were prioritized for causal claims, while qualitative and survey evidence was used to interpret

perceptions and mechanisms. K–12-only studies, purely technical papers without educational outcomes, unverified references, and retracted articles were excluded from the evidentiary conclusions. DOI metadata were cross-checked against publisher pages and Crossref. Eighty-four of the 85 studies in the principal supplementary list reported a DOI; the broader final reference set used in this paper was independently verified.

3.4 Data extraction

The public supplementary file was programmatically extracted into an effect-size-level file and a study-level file. The effect-size file retained all 387 coded records and included the SRL dimension, experimental and comparison sample sizes, feedback source, AI technology, message type, support type, modality, intervention length, academic subject, study design, and measurement method. The study-level file summarized 85 reports and included year, dominant technology, discipline, design, methodological quality, reference, and DOI.

The 387 rows were not treated as 387 independent studies. They are nested within 85 reports, and multiple outcomes can come from the same participants. Therefore, new analyses were limited to transparent descriptive counts. Statistical inference was taken from the published three-level models, which appropriately account for dependence among effect sizes (Huang et al., 2026).

3.5 Statistical analysis

Descriptive analyses included frequencies, percentages, publication-year trends, technology distributions, disciplinary distributions, design distributions, and the distribution of methodological-quality scores. For substantive effects, Hedges' g and its reported confidence or prediction intervals were extracted from published meta-analyses. Hedges' g is a standardized

mean difference adjusted for small-sample bias; positive values in the cited syntheses favor AI-supported learning. Moderator tests, multilevel meta-regression results, and heterogeneity measures were interpreted as originally reported. Estimates from different meta-analyses were displayed together for comparison but were **not pooled**, because they represent different populations, outcomes, comparators, and models.

3.6 Ethical and reproducibility considerations

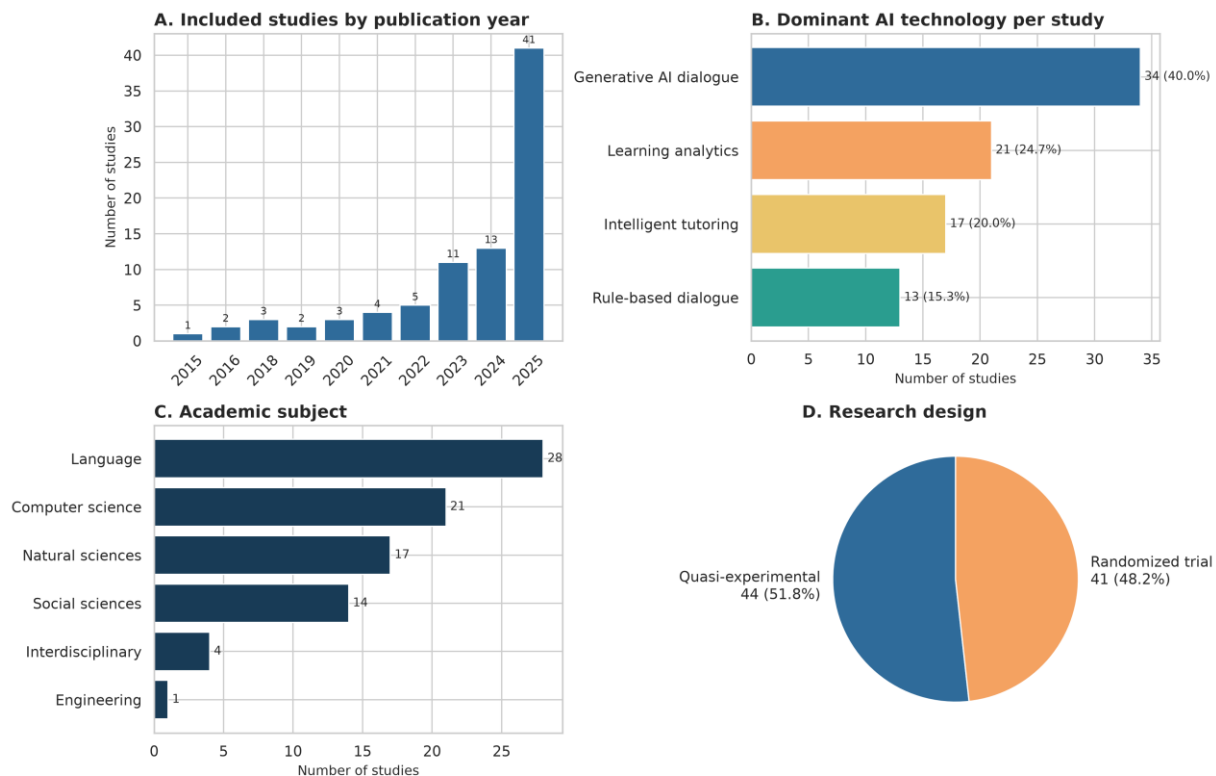
The study analyzed published aggregate data and did not involve direct participation by human subjects. No simulated or randomly generated observations were used. The reconstructed dataset, analysis scripts, descriptive tables, and figures accompany the research files so that counts and visualizations can be audited. The source meta-analysis remains the authority for its inferential estimates and deduplicated participant total.

4. Results

4.1 Profile of the quantitative evidence base

The 85-study corpus spans 2015–2025 and shows rapid growth after 2022. The annual counts were 1 study in 2015, 2 in 2016, 3 in 2018, 2 in 2019, 3 in 2020, 4 in 2021, 5 in 2022, 11 in 2023, 13 in 2024, and 41 in 2025. This concentration illustrates both the speed of research expansion and the immaturity of the evidence: much of the literature is recent and has had limited time for replication.

Profile of the 85-Study Higher-Education Evidence Base



Source: reconstructed from the published supplementary materials of Huang et al. (2026). Counts are study-level; no synthetic data.

Figure 1. The study-level evidence profile reconstructed from the supplementary corpus of Huang et al. (2026).

Generative-AI dialogue systems were the dominant technology in 34 studies (40.0%), followed by learning analytics in 21 (24.7%), intelligent tutoring systems in 17 (20.0%), and non-generative dialogue systems in 13 (15.3%). Seventy studies (82.4%) were coded as AI-led feedback and 15 (17.6%) as human–AI hybrid feedback. Text-only feedback dominated the corpus (57 studies, 67.1%), whereas 28 studies (32.9%) used multichannel feedback.

Study-level characteristic	Category	Studies	Percentage
AI technology	Generative-AI dialogue	34	40.0%
	Learning analytics	21	24.7%
	Intelligent tutoring	17	20.0%

	Rule-based dialogue	13	15.3%
Feedback source	AI-led	70	82.4%
	Human–AI hybrid	15	17.6%
Design	Quasi-experimental	44	51.8%
	Randomized trial	41	48.2%
Modality	Text only	57	67.1%
	Multichannel	28	32.9%

Language education was the largest disciplinary group with 28 studies (32.9%), followed by computer science with 21 (24.7%), natural sciences with 17 (20.0%), social sciences with 14 (16.5%), interdisciplinary contexts with 4 (4.7%), and engineering with 1 (1.2%). Intervention length was distributed across one session (25 studies, 29.4%), one to six weeks (18, 21.2%), six to ten weeks (26, 30.6%), and more than ten weeks (16, 18.8%). This disciplinary and temporal distribution limits generalization to all university contexts.

The mean methodological-quality score was 0.832 (SD = 0.114), the median was 0.83, and the interquartile range was 0.79–0.91. Although most studies clustered at relatively high scores, the range extended from 0.17 to 1.00.

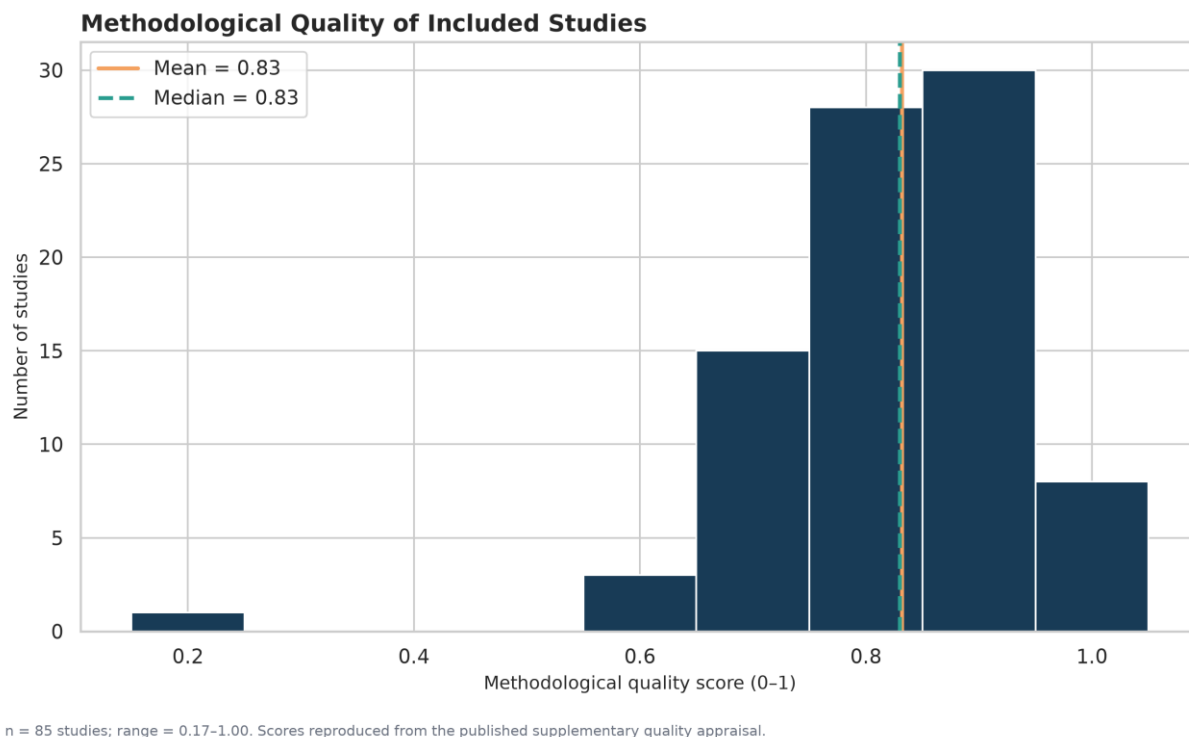


Figure 2. Distribution of methodological-quality scores reported for the 85 included studies.

4.2 Effects on teaching quality and learning outcomes

4.2.1 Academic performance

A meta-analysis of 19 studies and 24 effect sizes found that generative-AI interventions improved academic performance by **Hedges' $g = 0.683$** , 95% CI [0.17, 1.19], $p = .01$ (Gu & Yan, 2025). This is a meaningful average effect, but between-study inconsistency was very high, $I^2 = 94.4\%$, and the prediction interval ranged from -1.71 to 3.08 . Thus, the average benefit does not imply that every implementation improves achievement.

Teacher support was the most consequential moderator in that synthesis. Student–GenAI interaction supported by a teacher produced a large estimate, $g = 1.426$, whereas interaction without teacher support produced a near-zero estimate, $g = 0.077$ (Gu & Yan, 2025). Five of the 19 studies reported negative effects, and all five lacked teacher involvement. Although subgroup

results cannot by themselves prove causality, the pattern strongly supports pedagogical orchestration: students benefit when instructors structure the task, establish standards, guide prompting and verification, and connect AI output to disciplinary reasoning.

Primary studies provide compatible but more specific evidence. In the case of nursing training programs, the use of an AI chatbot that could be used on demand for domain-specific explanations significantly enhanced post-intervention knowledge among 73 participants ($p=.001$), highlighting the benefit of how immediate access to domain-specific explanations could be beneficial (Makhlouf et al., 2024). A knowledge-based virtual teaching assistant in Ghanaian higher education and an AI chatbot supporting college students' after-class review also reported improvements in learning-related outcomes, self-efficacy, attitudes, or motivation (Essel et al., 2022; Lee et al., 2022). Medical education and online computing quasi-experimental studies also suggest some benefits of chatbot programs and empathic feedback in conjunction with course-dependent activities for knowledge, motivation, self-regulation, and metacognitive reasoning (Han et al., 2022; Ortega-Ochoa et al., 2024). While these findings show promise for the use of targeted AI support, there is no justification for assuming any AI-generated content or output is automatically accurate or aligned with instructionally appropriate content.

4.2.2 Student engagement

The overall effect of ChatGPT-supported learning on students' overall engagement was medium, $g = 0.55$, with a significant amount of heterogeneity, $I^2 = 79\%$, across the 17 studies ($n = 1,735$) analyzed in this random-effects meta-analysis (Heung & Chiu, 2025). Behavioral engagement improved by $g = 0.454$, 95% CI [0.055, 0.853]; cognitive engagement improved by $g = 0.593$, 95% CI [0.378, 0.809]; and emotional engagement improved by $g = 0.512$, 95% CI [0.183, 0.842]. The higher cognitive estimate suggests that conversational AI can stimulate

strategy use, problem solving, explanation, and self-reflection when tasks require active dialogue rather than passive answer acceptance.

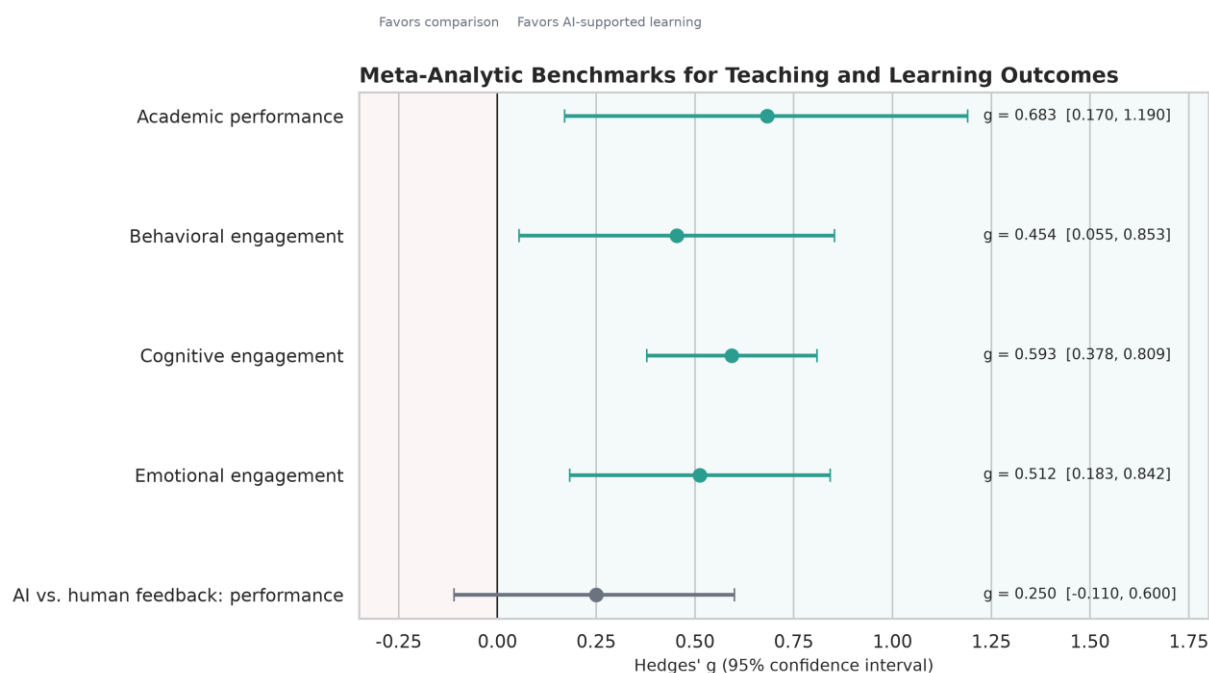


Figure 3. Benchmarks from individual syntheses in the form of a meta-analysis. There is no statistical combination of the rows. The interval between the AI and the human increases and fluctuates between positive and negative values (Gu & Yan, 2025; Heung & Chiu, 2025; Kaliisa et al., 2026).

Overall, students have settled positive but tempered views. The study surveyed 399 undergraduate and postgraduate students in Hong Kong and revealed that, along with the positive attitudes, there were concerns about accuracy, privacy, ethics, personal development, and career implications, when it comes to personalized support, writing and brainstorming support, and research related functions (Chan & Hu, 2023). A positive perceived usefulness can make for participation, but a positive perception does not equal learning. Institutions should thus use indicators of perception as well as performance, transfer, process and integrity.

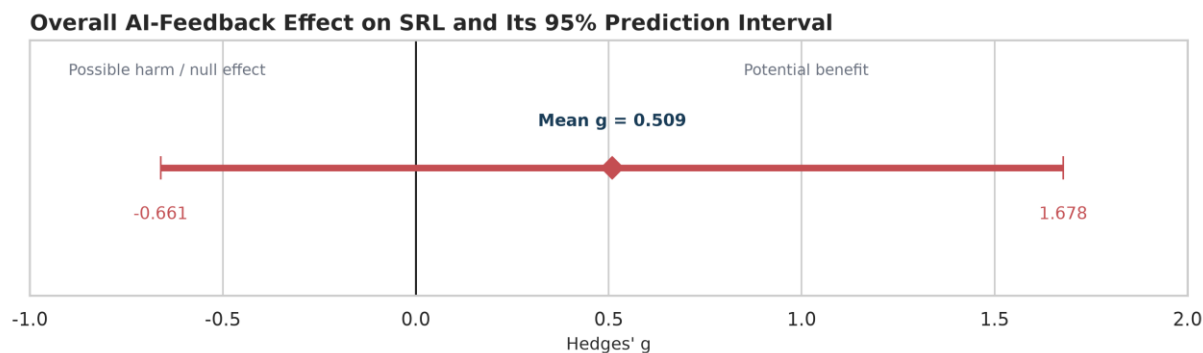
4.2.3 Quality of feedback and efficiency of teaching

Using AI to boost the quality of teaching by improving the timing, amount and frequency of feedback on student work. In the blended and flipped university courses, learning-analytics feedback has shown to be correlated with SRL, achievement, awareness, and recommendations based on student activities, which are beneficial for the students (Chen, 2024; Ustun et al., 2023). In the context of classroom chatbots, AI-based student assistants can be designed to provide supplementary support to the teacher's office hours and minimize teacher intervention, but still be accountable for course objectives and decisions regarding teacher intervention (Chen et al., 2023). Visual reports with AI support and personalized real-time scaffolds can facilitate students to gain insights into patterns of progress and strategy that are invisible (Liao et al., 2024; Lim et al., 2023).

But it's not proven that AI feedback is a better alternative to human feedback. In a meta-analysis of AI vs. human feedback, there was a small difference, $g = 0.25$, $[-0.11, 0.60]$ that contained a zero (Kaliisa et al., 2026). The right answer is that AI can provide access to feedback and can take over repetitive or iterative answers, but that human expert feedback is still needed to explain ambiguity, make disciplinary judgement, provide emotional support, ethical guidance and developmental interpretation.

4.3 Effects on self-directed and self-regulated learning

The three-level meta-analysis of 85 studies found a significant medium positive effect of AI feedback on overall SRL, **Hedges' $g = 0.509$** (Huang et al., 2026). The 95% prediction interval, $[-0.661, 1.678]$, is exceptionally important: a future implementation could plausibly yield a negative, null, moderate, or very large effect. The average is positive, but design and context determine whether this potential is realized.



The prediction interval, not a confidence interval, estimates the range in which a future study's true effect may fall. Its width shows that benefits are conditional.

Figure 4. The prediction interval is not for uncertainty about the mean, but for uncertainty about the future effects of the treatment (Huang et al., 2026).

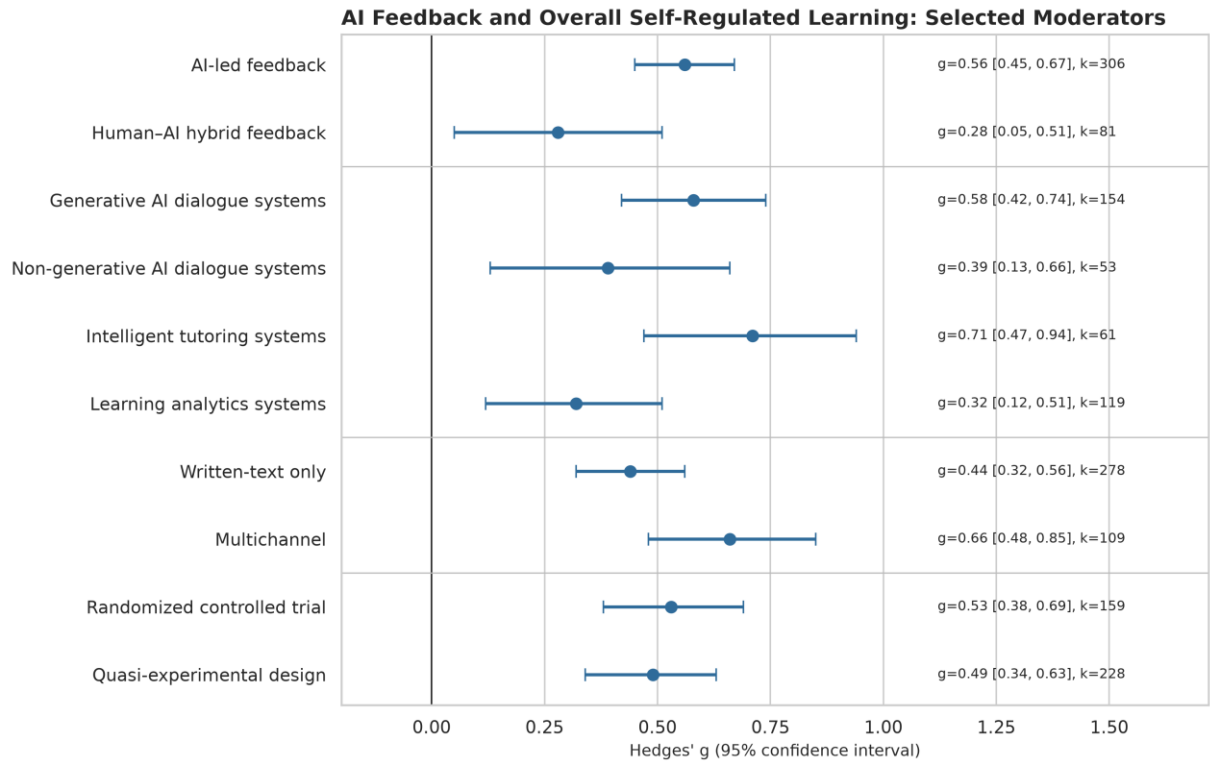
The average effect is explained by qualitative and mapping reviews. A systematic mapping review of 84 studies revealed that AI-SRL research is focused in higher education and that AI is applied to adaptive intervention/personalization, prediction and profiling, intelligent tutoring, assessment and evaluation. Cognitive and metacognitive regulation was investigated more frequently than motivational regulation and more than one-third of the studies failed to mention any SRL theory (Banihashem et al., 2025). A qualitative analysis of 14 higher-education studies identified evidence that supports the three stages of forethought, performance and reflection, and also noted a need to maintain human-centered agency while doing so (Lan & Zhou, 2025).

In a mixed methods study of 384 postsecondary writers and 10 interview participants, students often used ChatGPT as a brainstorming tool and for finding inspiration, felt responsible for a large part of their learning and were engaged in a variety of self-management strategies. The respondents to the survey, however, indicated lower levels of self-monitoring than; and participants differed in their opinion regarding their improvement in writing skills (Wang et al., 2024). The finding illustrates a recurring distinction: AI can increase access, practice, and task motivation without automatically producing accurate self-evaluation or durable skill growth.

4.3.1 Overall moderators

AI-led feedback produced $g = 0.56$, 95% CI [0.45, 0.67], whereas human–AI hybrid feedback produced $g = 0.28$, 95% CI [0.05, 0.51]. The difference was statistically significant, $F(1) = 4.81$, $p = .029$ (Huang et al., 2026). This result does **not** demonstrate that instructors should be removed. The categories describe study implementations, not randomized assignment to “teacher present” versus “teacher absent” across identical contexts. Hybrid systems may have been used with more difficult tasks or learners, and the academic-performance meta-analysis reached the complementary conclusion that teacher scaffolding greatly strengthened GenAI effects. Taken together, the evidence supports clear role allocation rather than teacher replacement.

Intelligent tutoring systems produced the largest technology subgroup estimate, $g = 0.71$, 95% CI [0.47, 0.94], followed by generative-AI dialogue systems, $g = 0.58$ [0.42, 0.74], non-generative dialogue systems, $g = 0.39$ [0.13, 0.66], and learning-analytics systems, $g = 0.32$ [0.12, 0.51]. The omnibus technology difference was marginal, $F(3) = 2.58$, $p = .054$. Multichannel feedback, $g = 0.66$ [0.48, 0.85], exceeded text-only feedback, $g = 0.44$ [0.32, 0.56], but the difference was again marginal, $F(1) = 3.82$, $p = .051$. These estimates suggest promising design directions, not definitive technology rankings.



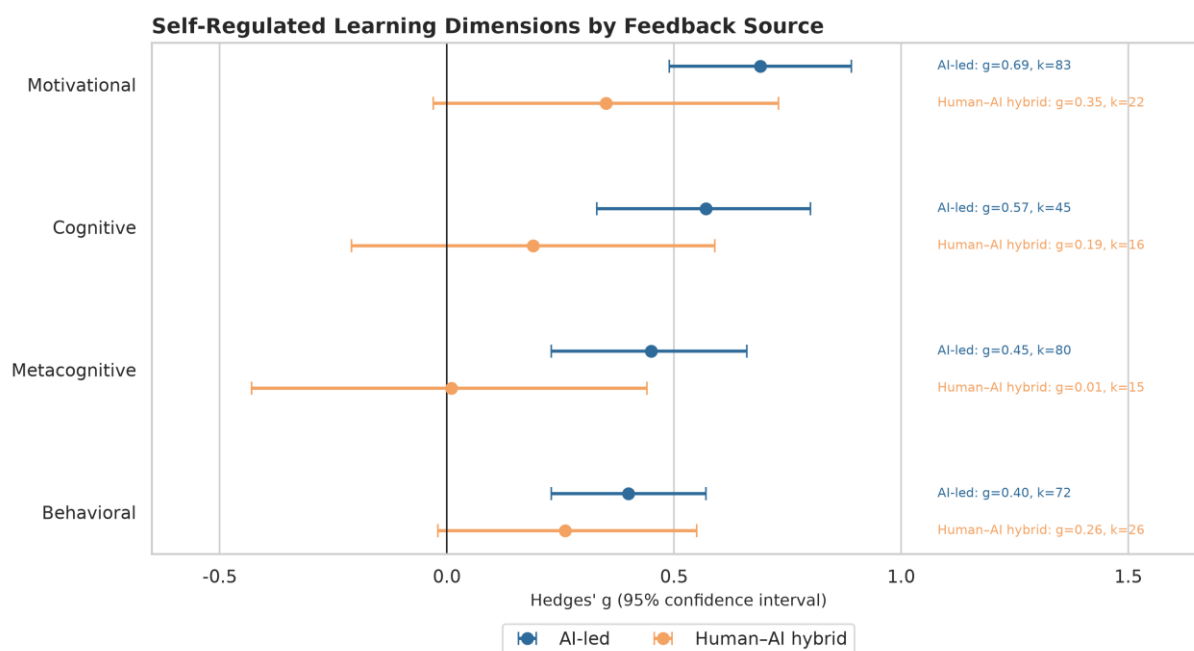
Source: Huang et al. (2026), Supplementary Table I1. k denotes dependent effect-size estimates, not independent studies.

Figure 5. Estimates for overall SLR from selected moderators. The *k* in this case represents dependent estimates of effect size instead of independent studies (Huang et al., 2026).

A hierarchical multilevel meta-regression strengthened the interpretation. Adding feedback source and technology type significantly improved fit relative to the contextual model, likelihood-ratio $\chi^2(4) = 13.08$, $p < .05$. Across expanded models, human–AI hybrid feedback retained a negative coefficient relative to AI-led feedback, approximately $b = -0.32$ to -0.36 , $p < .05$. Adding message type, support type, or measurement method did not provide comparable improvement. These associations identify agent configuration and technology class as major sources of heterogeneity but do not establish simple causal superiority.

4.3.2 Cognitive, metacognitive, motivational, and behavioral dimensions

Within the AI-led subgroup, the largest dimension-specific estimate was motivational, $g = 0.69$, 95% CI [0.49, 0.89], followed by cognitive, $g = 0.57$ [0.33, 0.80], metacognitive, $g = 0.45$ [0.23, 0.66], and behavioral, $g = 0.40$ [0.23, 0.57]. Human–AI hybrid estimates were smaller, and their intervals crossed zero in all four dimensions. This pattern is consistent with the conclusion that AI feedback more readily influences motivation and cognitive strategy use than sustained behavior and metacognitive control.



Source: Huang et al. (2026), Supplementary Tables I2–I5. Hybrid estimates for cognitive, metacognitive, motivational, and behavioral dimensions cross zero.

Figure 6. Dimension-specific subgroup estimates by feedback source. Human–AI hybrid confidence intervals cross zero for all four dimensions (Huang et al., 2026).

The motivational outcome seems inviting since the time spent waiting can be minimized, practice can be nonjudgmental, examples can be created, and competence can be enhanced. The cognitive outcome is also likely as students can ask questions about explanations, look at

alternatives and is able to get problem specific prompts. Metacognition and behavior require more than access to information: learners must monitor the reliability of their own understanding, resist premature closure, manage time and effort, and make decisions independently. These are precisely the processes most vulnerable to over-assistance.

A significant design difference appeared in behavioral outcomes. Randomized trials yielded $g = 0.52$, 95% CI [0.31, 0.73], whereas quasi-experimental studies yielded $g = 0.21$ [0.00, 0.41], $F(1) = 4.47$, $p = .037$. Academic subject was marginally related to metacognitive effects, $F(5) = 2.30$, $p = .051$. Given the number of subgroup comparisons and unequal cell sizes, these secondary findings should be interpreted as hypotheses for future research rather than stable laws.

4.4 Risks and boundary conditions

4.4.1 Cognitive offloading and shallow learning

The strongest caution comes from studies that separate immediate task quality from later learning. In a randomized experiment with 117 university students, the ChatGPT group improved essay scores more than comparison groups, but knowledge gain and transfer did not differ significantly. Process data showed that there was some dependency and engagement in monitoring and evaluation was lowered, which has been called “metacognitive laziness” (Fan et al., 2025). A different experiment reported that using LLM lead to lowered mental load and shallowness of the student's scientific inquiry (Stadler et al., 2024).

The results do not necessarily imply that reducing efforts is always bad. Instructing should eliminate unnecessary effort and still include germane effort (reasoning, retrieval, evaluation, relationships are the way in which learning occurs, including relationship between

and with others. If AI eliminates germane effort, students may become faster without becoming more capable.

4.4.2 Reliability, verification, and feedback dependence

Generative output can be fluent but inaccurate. Review studies highlight the importance of pedagogically relevant feedback, actionable feedback, and interpretable feedback that is accurate and trusted by students (Ba et al., 2025; Chan & Hu, 2023). Students who cannot evaluate evidence may accept false explanations or fabricated references. Verification must therefore be taught as a learning outcome: students should trace claims to course texts and scholarly sources, compare alternatives, identify uncertainty, and document how AI output was checked.

4.4.3 Academic integrity and assessment validity

When AI can produce a satisfactory response to an assessment without the learner performing the intended cognition, the problem is not only misconduct; it is a validity problem. Universities must clarify permitted assistance and redesign assessments to make reasoning visible. Process evidence, oral defense, staged drafts, annotated decisions, local data, reflective commentary, and supervised performance can distinguish support from substitution. Assessment should include moments of independent performance so that students and instructors can determine what has actually been learned.

4.4.4 Privacy, bias, and equity

Learning analytics and adaptive systems can process detailed behavioral, academic, and sometimes affective data. A meta-systematic review revealed that the reviews of AI in higher education lacked consistency in ethics, transparency, contextual reporting, collaboration, and

methodological rigor (Bond et al., 2024). Data minimization, informed consent, transparent purpose limitation, accessibility, human review of high-stakes decisions, and evaluation across student groups are therefore essential components of teaching quality rather than external compliance concerns.

4.5 Summary of statistical findings

Outcome or comparison	Estimate	Uncertainty or heterogeneity	Interpretation
Overall AI feedback → SRL	$g = 0.509$	95% prediction interval $[-0.661, 1.678]$	Medium mean effect; effects are highly conditional.
AI-led feedback → overall SRL	$g = 0.56$	95% CI $[0.45, 0.67]$	Moderate positive subgroup effect.
Human–AI hybrid feedback → overall SRL	$g = 0.28$	95% CI $[0.05, 0.51]$	Smaller positive subgroup effect.
GenAI → academic performance	$g = 0.683$	95% CI $[0.17, 1.19]$; $I^2 = 94.4\%$	Positive mean effect with very high heterogeneity.
Teacher-supported GenAI → performance	$g = 1.426$	Moderator estimate	Large benefit under teacher scaffolding.
GenAI without teacher support → performance	$g = 0.077$	Moderator estimate	Near-zero benefit without teacher support.
ChatGPT → overall engagement	$g = 0.55$	$I^2 = 79\%$	Medium positive effect with substantial heterogeneity.
AI versus human feedback → performance	$g = 0.25$	95% CI $[-0.11, 0.60]$	No clear superiority over human feedback.

5. Discussion

5.1 The central finding: positive on average, conditional in practice

The evidence supports a balanced conclusion. AI applications can improve university teaching quality by increasing access to explanations, enabling repeated practice, accelerating formative feedback, supporting personalization, and giving instructors and students greater visibility into learning processes. They can also strengthen motivational and cognitive components of self-directed learning. However, the wide prediction intervals and substantial heterogeneity show that “AI use” is not a single intervention. Outcomes depend on technology, task, discipline, learner readiness, feedback design, teacher involvement, duration, and the degree to which students remain cognitively responsible.

The average-effect interpretation must therefore be stated carefully. A medium pooled effect is not a guarantee for a new course. It is a summary of diverse implementations. University leaders should not translate the evidence into a general mandate to “use AI.” They should identify the instructional problem, select an AI function suited to that problem, define what students must continue to do independently, and evaluate both intended benefits and possible displacement of learning.

5.2 Reconciling AI-led feedback with teacher-supported AI

At first glance, two findings appear contradictory. In the SRL meta-analysis, AI-led feedback produced a larger subgroup estimate than human–AI hybrid feedback. In the academic-performance meta-analysis, teacher-supported GenAI greatly outperformed unsupported GenAI use. The apparent contradiction disappears when the comparisons are examined closely.

The SRL categories aggregate different studies and outcomes; they do not randomly assign identical learners and tasks to the presence or absence of an instructor. AI-led systems may deliver more consistent or intensive prompts, whereas hybrid systems may appear in complex settings where human support is added because learners need it. The academic-performance moderator directly addresses teacher scaffolding during student-GenAI interaction. The defensible synthesis is that AI should be able to deliver immediate routine feedback, but teachers must design the learning environment, establish standards, support interpretation, and intervene when judgment or motivation requires human expertise.

5.3 Why motivation and cognition improve more readily than metacognition

AI's conversational and adaptive affordances provide rapid experiences of progress. Students can obtain an explanation, example, translation, critique, or coding hint at the point of need. This responsiveness can enhance competence beliefs, reduce anxiety, and sustain effort. It can also be used to engage students' thinking if they are asking for clarification or exploring options.

Metacognition will only be developed if students monitor what they are thinking and make strategic changes. A system that supplies conclusions too quickly can suppress this cycle. Consequently, instructional design should make AI use reflective by requirement, not reflective by hope. Students can be required to predict before querying, explain why they accepted or rejected suggestions, identify uncertainty, compare AI output with a scholarly source, and complete an unaided transfer task. These activities convert AI from an answer generator into a metacognitive scaffold.

5.4 Implications for teaching quality

High-quality AI-supported teaching is not defined by technological novelty. It is defined by alignment among outcomes, activities, feedback, assessment, and learner support. AI can improve quality where it solves real constraints such as delayed feedback, insufficient practice, inaccessible explanations, or limited visibility into progress. It can lower quality when it introduces generic content, weakens teacher–student dialogue, masks misconceptions, or encourages students to optimize products rather than understanding.

Student feedback from the review revealed that students appreciate individualized support, brainstorming, but are worried about issues of reliability, privacy, ethics and personal development (Chan & Hu, 2023). Explicit AI literacy is thus part of the teaching: recognizing the limitations of the models, advising responsibly and correctly, checking the validity of the models, safeguarding sensitive data, citing the assistance according to institutional rules, and recognizing when it is necessary to consult a human being.

5.5 AI model is teacher-led

Structured prompts that guide students to use AI tools for higher-order thinking, knowledge building and self-regulation (such as guided ChatGPT tasks and reflective-thinking mechanisms) are designed to match the students' use of the AI tool and do not allow students to interact with an untargeted tool (Lee et al., 2024; Liu et al., 2023).

The evidence supports a four-stage implementation model.

Stage	Instructor role	Student role	AI role	Evidence of learning
1. Model	Applies disciplinary logic, criteria, verification and ethics.	Notifies, asks questions and makes comparisons of examples.	Gives contrasting outputs/examples.	Student is able to point out strengths, errors and criteria.
2. Scaffold	Initiates, guides and reviews	Takes part in the task, asks	Gives suggestions,	Note revision process and

	routines, prompts and checkpoints.	questions of AI and explains decisions.	clarifications and diagnostic feedbacks	feedback are clearly evident.
3. Fade	Minimizes prompts and increases the level of difficulty.	Uses strategies and checks work for self.	Provides some or optional assistance.	Student uses strategies and identifies his/her errors with less assistance.
4. Transfer	Assesses independent application and reflection.	Performs a novel task without AI, then reflects on tool use.	Absent during the core performance or used only after completion.	Unaided performance, transfer, and calibrated self-evaluation.

This model addresses the central risk revealed by the evidence: short-term improvement can occur without durable learning. The stages of fading and transfer allow for student-owned competence to be created with the support of AI.

6. Suggestions for Universities and Instructors

6.1 Link the use of Aligned AI to clear learning outcomes

Each activity, with the use of AI, should be able to identify the intended learning process. When the result is argumentation, students will need to make and support arguments, rather than simply edit a generated text. When the result is programming, AI suggestions should aid in debugging, but not take the place of algorithmic thinking. If the outcome is clinical judgment, AI can generate cases or feedback, but learners must justify decisions against evidence and professional standards.

6.2 Preserve productive struggle

Instructors should distinguish assistance that removes irrelevant barriers from assistance that removes the target cognition. A useful pattern is **attempt–query–evaluate–revise**: students

first attempt the task, then consult AI, evaluate the response against criteria and sources, and revise with an explanation. This sequence protects retrieval, diagnosis, and monitoring.

6.3 Foster knowledge of AI and feedback literacy.

There is a need for systematic practice of writing task specific prompts, identifying uncertainty, checking citations, testing alternate explanations, and determining if feedback is task appropriate. They should document verification and explain which suggestions they rejected. Such practices strengthen evaluative judgment and reduce blind acceptance.

6.4 Redesign assessment for process and transfer

Assessment should include staged submissions, prompt or interaction logs where appropriate, annotated revisions, oral questioning, personalized or local applications, and unaided components. Policies should specify permitted use by task rather than relying only on broad prohibitions. The goal is to preserve assessment validity while recognizing legitimate learning support.

6.5 Combine automated and human feedback strategically

AI is well suited to frequent low-stakes feedback, routine error detection, practice generation, progress summaries, and first-pass explanations. This human feedback should focus on: nuances of the discipline, conflicting evidence, misunderstandings, emotional support, ethical judgment, creativity, developmental priorities. This division can be made without the perception of human and AI feedback being interchangeable.

6.6 Protect privacy, equity, and accessibility

Universities should approve tools through data-protection and accessibility review, minimize the use of identifiable student data, publish clear retention and model-training policies,

and provide alternatives for students who cannot or should not use a particular system. Outcomes should be disaggregated to detect whether language background, disability, prior knowledge, or digital access affects benefit.

6.7 Hire faculty for their capacity to build the school's capacity.

Pedagogy, assessment, disciplinary use, ethics, and evaluation as well as tool operation should be included in faculty development. Time and institutional support is required by instructors for testing outputs, creating examples, setting up course policies and tracking effects. Without professional learning, there is likely to be more inconsistency in implementation of AI.

6.8 Assess various aspects of quality

Don't measure the success of AI programs solely by usage and student satisfaction. A balanced evaluation should include academic performance, transfer, higher-order thinking, SRL, engagement, feedback uptake, equity, integrity, workload, privacy incidents, and student ability to perform without AI.

Evaluation domain	Example indicator	Why it matters
Learning	Achievement and transfer task performance	Distinguishes product improvement from durable learning.
SRL/SDL	Goal quality, monitoring accuracy, strategy adaptation, help seeking	Measures growth in learner agency.
Engagement	The factors of participation, persistence, and cognitive investment, as well as emotional response, are discussed.	Determines the level of sustained meaningful involvement that is provided by AI.
Feedback	Uptake, quality of revisions and explaining changes	Determines if feedback is given or accepted.
Teaching	Rewards and incentives, instructor workload, quality of intervention, student support	The assessment of the role AI plays in enhancing teaching and learning abilities.
Equity	Inequalities and inequity in outcomes and access	Identifies inequity or inequality.
Integrity and ethics	The disclosure and verification, privacy and incident data.	Ensures the validity and integrity of assessment and trust.

7. Limitations

This study has several limitations. First, the principal quantitative dataset is a secondary reconstruction of published supplementary material; the inferential models were not independently re-estimated because the file provided coded characteristics and moderator results rather than a complete raw effect-size calculation dataset. Second, the field is rapidly changing. Forty-one of the 85 principal studies were published in 2025, so replication and longer follow-up are limited. Third, the evidence is concentrated in language learning, computer science, and natural sciences, with minimal engineering representation. Fourth, outcomes and comparators differ across syntheses, preventing a valid single pooled estimate for “teaching quality.” Fifth, many interventions are short, and immediate performance can overstate durable learning. Sixth, subgroup and moderator findings are observational across studies and can be confounded by task difficulty, learner characteristics, implementation intensity, and study quality. Finally, English-language and DOI-verifiable research was prioritized, which increases traceability but may omit relevant local or non-English evidence.

8. Conclusion

AI applications can improve the quality of university teaching and develop important components of students’ self-directed learning. The strongest benefits arise from timely feedback, personalized explanations, repeated practice, learning visibility, and structured dialogue. Quantitative evidence indicates medium positive average effects on SRL, academic performance, and engagement. However, these averages conceal substantial variation, and AI feedback does not consistently outperform expert human feedback.

The most important conclusion is pedagogical rather than technological. **AI improves learning when it augments a coherent instructional design; it weakens learning when it replaces the reasoning, monitoring, dialogue, or independent performance that the course intends to develop.** Universities should therefore adopt teacher-orchestrated AI, explicitly teach verification and feedback literacy, preserve productive struggle, fade support, assess transfer, and evaluate equity and ethics alongside achievement. Under these conditions, AI can become a meaningful scaffold for higher-quality teaching and increasingly autonomous learners.

References

1. Ba, S., Yang, L., Yan, Z., Looi, C. K., & Gašević, D. (2025). Unraveling the mechanisms and effectiveness of AI-assisted feedback in education: A systematic literature review. *Computers and Education Open*, 9, Article 100284. <https://doi.org/10.1016/j.caeo.2025.100284>
2. Banihashem, S. K., Bond, M., Bergdahl, N., Khosravi, H., & Noroozi, O. (2025). A systematic mapping review at the intersection of artificial intelligence and self-regulated learning. *International Journal of Educational Technology in Higher Education*, 22(1), Article 50. <https://doi.org/10.1186/s41239-025-00548-8>
3. Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21(1), Article 4. <https://doi.org/10.1186/s41239-023-00436-z>
4. Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies and academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, 27, 1–13. <https://doi.org/10.1016/j.iheduc.2015.04.007>
5. Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational*

- Technology in Higher Education, 20(1), Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
6. Chen, J. (2024). Effects of learning analytics-based feedback on students' self-regulated learning and academic achievement in a blended EFL course. *System*, 124, Article 103388. <https://doi.org/10.1016/j.system.2024.103388>
 7. Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers*, 25(1), 161–182. <https://doi.org/10.1007/s10796-022-10291-4>
 8. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), Article 22. <https://doi.org/10.1186/s41239-023-00392-8>
 9. Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
 10. Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerd, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, 162, Article 104094. <https://doi.org/10.1016/j.compedu.2020.104094>
 11. Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *International Journal of Educational Technology in Higher Education*, 19(1), Article 57. <https://doi.org/10.1186/s41239-022-00362-6>
 12. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>
 13. Gu, J., & Yan, Z. (2025). Effects of GenAI interventions on student academic performance: A meta-analysis. *Journal of Educational Computing Research*, 63(6), 1460–1492. <https://doi.org/10.1177/07356331251349620>

14. Guan, R., Raković, M., Chen, G., & Gašević, D. (2025). How educational chatbots support self-regulated learning? A systematic review of the literature. *Education and Information Technologies*, 30(4), 4493–4518. <https://doi.org/10.1007/s10639-024-12881-y>
15. Han, J.-W., Park, J., & Lee, H. (2022). Analysis of the effect of an artificial intelligence chatbot educational program on non-face-to-face classes: A quasi-experimental study. *BMC Medical Education*, 22(1), Article 830. <https://doi.org/10.1186/s12909-022-03898-3>
16. Heung, Y. M. E., & Chiu, T. K. F. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence*, 8, Article 100361. <https://doi.org/10.1016/j.caeai.2025.100361>
17. Huang, C.-Q., Lu, L.-N., Huang, Q.-H., Zhang, Y.-R., He, T., Tu, Y.-F., & Hwang, G.-J. (2026). Effects of artificial intelligence feedback on students' self-regulated learning in higher education: A three-level meta-analysis. *Educational Psychology Review*, 38, Article 64. <https://doi.org/10.1007/s10648-026-10166-z>
18. Jin, S.-H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, 20(1), Article 37. <https://doi.org/10.1186/s41239-023-00406-5>
19. Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2026). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educational Psychology*, 46(1), 80–111. <https://doi.org/10.1080/01443410.2025.2553639>
20. Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *npj Science of Learning*, 10(1), Article 21. <https://doi.org/10.1038/s41539-025-00319-0>
21. Lee, H.-Y., Chen, P.-H., Wang, W.-S., Huang, Y.-M., & Wu, T.-T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education*, 21(1), Article 16. <https://doi.org/10.1186/s41239-024-00447-4>

22. Lee, Y.-F., Hwang, G.-J., & Chen, P.-Y. (2022). Impacts of an AI-based chatbot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. *Educational Technology Research and Development*, 70(5), 1843–1865. <https://doi.org/10.1007/s11423-022-10142-8>
23. Liao, X., Zhang, X., Wang, Z., & Luo, H. (2024). Design and implementation of an AI-enabled visual report tool as formative assessment to promote learning achievement and self-regulated learning: An experimental study. *British Journal of Educational Technology*, 55(3), 1253–1276. <https://doi.org/10.1111/bjet.13424>
24. Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, 139, Article 107547. <https://doi.org/10.1016/j.chb.2022.107547>
25. Liu, C., Hou, J., Tu, Y.-F., Wang, Y., & Hwang, G.-J. (2023). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, 31(9), 5614–5632. <https://doi.org/10.1080/10494820.2021.2012812>
26. Makhoulouf, E., Alenezi, A., & Shokr, E. A. (2024). Effectiveness of designing a knowledge-based artificial intelligence chatbot system into a nursing training program: A quasi-experimental design. *Nurse Education Today*, 137, Article 106159. <https://doi.org/10.1016/j.nedt.2024.106159>
27. Ortega-Ochoa, E., Quiroga Pérez, J., Arguedas, M., Daradoumis, T., & Marquès Puig, J. M. (2024). The effectiveness of empathic chatbot feedback for developing computer competencies, motivation, self-regulation, and metacognitive reasoning in online higher education. *Internet of Things*, 25, Article 101101. <https://doi.org/10.1016/j.iot.2024.101101>
28. Page, M. J., Moher, D., 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., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration:

Updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160.

<https://doi.org/10.1136/bmj.n160>

29. Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, Article 422.

<https://doi.org/10.3389/fpsyg.2017.00422>

30. Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>

31. Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 160, Article 108386. <https://doi.org/10.1016/j.chb.2024.108386>

32. Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology*, 66, Article 101976.

<https://doi.org/10.1016/j.cedpsych.2021.101976>

33. Ustun, A. B., Zhang, K., Karaoğlu-Yılmaz, F. G., & Yılmaz, R. (2023). Learning analytics based feedback and recommendations in flipped classrooms: An experimental study in higher education. *Journal of Research on Technology in Education*, 55(5), 841–857. <https://doi.org/10.1080/15391523.2022.2040401>

34. Wang, C., Li, Z., & Bonk, C. J. (2024). Understanding self-directed learning in AI-assisted writing: A mixed methods study of postsecondary learners. *Computers and Education: Artificial Intelligence*, 6, Article 100247.

<https://doi.org/10.1016/j.caeai.2024.100247>

35. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), Article 39. <https://doi.org/10.1186/s41239-019-0171-0>