ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5691
https://doi.org/10.24059/olj.v30iSI2.5691

A Meta-Analysis of AI-Assisted Self-Regulated Learning

Zhuo Zhang
Towson University, USA
Belle Li
Purdue University, USA
Qian Xu
Iowa State University, USA
Zui Cheng
Shenzhen University, China
Richard E. West
Brigham Young University, USA

Abstract

This meta-analysis examines the effects of artificial intelligence (AI)-based pedagogical interventions on learners' self-regulated learning and its subdomains, including cognitive, metacognitive, and resource management strategies. In addition, the study investigates whether these effects vary across four groups of moderators concerning research context, intervention design, AI, and self-regulated learning. A systematic search of databases identified an initial pool of 5,741 potentially relevant studies, of which 34 met all inclusion criteria. These studies yielded a total of 91 effect sizes. Results showed a positive overall effect of AI-based pedagogical interventions on self-regulated learning (Hedges' g = 0.76). Positive effects were also observed across SRL subdomains (g = 0.48 for cognitive strategies, 0.73 for metacognitive strategies, and 0.93 for resource management strategies). Moderator analyses showed that the effects of AI-based interventions on self-regulated learning varied significantly across educational levels, AI pedagogical roles, and types of AI applications, whereas other factors such as discipline, intervention length, type of control group, sample size, SRL measure did not show significant differences. By synthesizing existing empirical evidence, this meta-analysis provides a quantitative foundation for understanding how and under what conditions AI can be effectively integrated into education to support learners' self-regulated learning.

Keywords: self-regulated learning, artificial intelligence, meta-analysis

Introduction

Self-regulated learners take an active role in planning, monitoring, managing, and evaluating their learning processes and outcomes to attain predetermined goals (Pintrich, 2000, 2004). Self-regulated learning (SRL) is therefore widely recognized as a key competency associated with academic achievement, motivation, and satisfaction across various educational contexts, including K-12 education (Dent & Koenka, 2016), higher education (Broadbent & Poon, 2015; Cheng et al., 2025), and continuing education (Hemmler & Ifenthaler, 2024).

Given its importance across multiple dimensions of learning outcomes, a substantial body of research has examined interventions designed to strengthen learners' SRL capacity, particularly through the integration of educational technology. Among these technologies, artificial intelligence (AI) has emerged as an increasingly prominent approach. In recent years, researchers have shown growing interest in incorporating AI technologies, such as large language models (LLMs), chatbots, intelligent tutoring systems, and learning analytics dashboards, into pedagogical interventions aimed at supporting learners' SRL (e.g., Fleur et al., 2023; Liu et al., 2024; McCarthy et al., 2018; Ortega-Ochoa et al., 2024).

Corresponding author: Zhuo Zhang, zhuozhang@towson.edu
This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).
© 2026 The Author(s)

The growing body of empirical evidence on AI-based pedagogical interventions (AI-PI) and SRL necessitates a rigorous quantitative research synthesis to address two key questions. First, do AI-PI enhance learners' SRL, and if so, to what extent? Second, what factors, such as AI's pedagogical role, AI tool type, learners' educational level, and disciplinary context, may influence the effects of AI-PI on SRL? Existing quantitative research syntheses on AI integration in education have largely focused on learning achievement or outcomes (e.g., Dai et al., 2024; Deng et al., 2025; Dong et al., 2025; Zheng et al., 2023). In contrast, the ways in which AI integration shapes how learners learn (their learning strategies and behaviors) have received far less systematic attention.

To address this gap, the present research synthesis integrates existing quantitative evidence on AI-PI and SRL. By adopting meta-analytic procedures, this study aims to provide a comprehensive assessment of the overall effects of AI-PI on SRL and to examine the conditions under which these effects vary.

Literature Review

Self-Regulated Learning (SRL) and SRL Strategies

Self-regulated learning (SRL) refers to a purposeful process in which learners set goals and actively monitor, adjust, and manage their cognition, motivation, and behaviors (Pintrich, 2000). Drawing on Weinstein et al.'s (2000) definition of learning strategies as "any thoughts, behaviors, beliefs, or emotions that facilitate the acquisition, understanding, or later transfer of new knowledge and skills" (p. 727), this study defines SRL strategies as the intentional thoughts, behaviors, beliefs, and emotional actions learners use to set goals and to regulate their cognition, motivation, and behaviors in support of learning and knowledge transfer.

SRL strategies can be further categorized by learning domain and phase. A common domain-based distinction includes metacognitive, cognitive, and resource management learning strategies (Hemmler & Ifenthaler, 2024; Pintrich et al., 1999; Theobald, 2021). According to Pintrich et al. (1999), cognitive learning strategies are strategies that help learners both remember information and comprehend more complex content, such as rehearsal, elaboration, and organization. Metacognitive strategies are strategies that support planning, monitoring, and reflecting activities, such as goal setting, task analysis, and self-evaluation. Resource management strategies are strategies that help manage and control internal and external resources such as effort, time, learning environment, and interpersonal support. Examples include effort regulation, time management, environmental structuring, and help-seeking. A widely adopted phase-based distinction is Zimmerman's (2000) cyclical three-phase model, which includes the forethought, performance, and self-reflection phases. In forethought phase, sample learning strategies include goal setting, planning, and activation of prior knowledge; in performance phase, sample learning strategies include help-seeking, time management, environmental structuring, rehearsal, and elaboration; and in the reflection phase, sample learning strategies include self-reflection and self-assessment.

Artificial Intelligence in Education

Education, like nearly all disciplines, has been affected by the emergence of generative artificial intelligence tools. While artificial intelligence tools have been studied since the 1950s with early focus on natural language processing and machine translation to generate simple sentences and paragraphs (Yu & Guo, 2023), more recent generative versions of these technologies produce output that is more interactive, more helpful, and more accurate (while deep understanding of concepts as yet is still elusive). These tools can impact education including supporting teachers in creating lesson plans (Gunbater et al., 2025), summarizing discussion comments to facilitate focused interactions (Bauer et

al., 2023), and providing more detailed feedback to students (West & Wiley, 2024). In addition, AI tools provide many opportunities to students, including support in understanding concepts through chatbot interactions (Thottoli et al., 2024), scaffolding in writing (Kim & Kim, 2022), and even making educational materials more accessible and personalized (Wiley, 2025). In addition, AI tools have been discussed as problematic for education in that it may facilitate cheating (Elom et al., 2025), limit human interactions (Alwaqdani, 2025), or impact students negatively due to bias (Warr et al., 2024).

Perhaps as would be expected, much of the discussion around the use of AI in education has lacked nuance, either espousing its great benefits or significant harms. For example, Nedungadi et al. (2024) argued that Generative AI provided "transformative power" for achieving quality education that is equitable and lifelong. Konstantinova et al. (2023) similarly argued that Generative AI has "all the potential" for solving educational challenges. In addition, much of the early research on AI in education has focused on teacher readiness or acceptance (Alwakdani, 2025; Yue et al., 2024) and AI literacy (Su et al., 2023). However, some scholars are calling for AI scholarship that delves deeper into how AI can support various kinds of learning. For example, Ouyang and Jiao (2021) argued that AI research could be organized into three paradigms: AI-directed, AI-supported, and AI-empowered learning. This final paradigm represents the "learner as leader" where learners work with AI to take agency over their learning in powerful ways. Part of this empowerment could be using AI to improve self-regulated learning, or the empowerment of the individual to best control their own learning processes.

AI-Based Pedagogical Interventions (AI-PI) and Self-Regulated Learning

In this study, AI-PI are defined as intentional implementations of AI-powered tools and systems designed to support, scaffold, or enhance students' learning processes and outcomes. Recent systematic reviews have documented the growing body of research on how various AI technologies can be integrated into educational practice to support learning (Banihashem et al., 2025; Lan & Zhou, 2025), including scaffolding students' SRL processes. AI-driven educational tools, such as intelligent tutoring systems, adaptive learning platforms, and conversational agents, are increasingly used to provide personalized guidance and feedback that support key SRL processes, including goal setting, strategic planning, monitoring, and reflection (Guan et al., 2025; Lan & Zhou, 2025). For example, chatbot tutors can prompt learners to identify resources, apply appropriate learning strategies, and monitor their understanding in real time.

Empirical studies across different educational contexts have reported improvements in students' SRL when AI tools are integrated into learning activities. In the domain of language learning, for instance, Wei (2023) showed that college-level English learners who engaged with an AI-powered learning platform exhibited significantly greater gains in SRL strategies compared to a control group receiving traditional instruction. The AI-mediated instruction in Wei's study provided interactive exercises, instant feedback, and adaptive hints, which helped students plan their study, monitor progress, and adjust strategies more effectively than their peers in the non-AI condition (Wei, 2023). Similarly, in science education, Ng et al. (2024) conducted a pilot experiment comparing a generative AI-based chatbot to a rule-based chatbot for supporting high school students' SRL during a three-week online course. Their findings indicated that the AI-enhanced chatbot ("SRLbot," powered by ChatGPT) led to higher levels of student engagement, more frequent help-seeking and self-monitoring behaviors, and improved science learning outcomes, relative to the traditional rule-based tutor (Ng et al., 2024). Students using the generative AI tutor reported that its personalized recommendations and adaptive responses helped them maintain regular study habits and reduce learning anxiety, thereby boosting motivation and proactive SRL behaviors (Ng et al., 2024). These examples align with a broader trend in the literature: well-designed AI interventions (e.g. intelligent tutors, recommender systems, and adaptive feedback agents) can effectively mimic the one-on-one support of a human coach, prompting learners to utilize SRL strategies more consistently and confidently (Chang & Sun, 2024).

Traditional and AI-Based Pedagogical Interventions for Self-Regulated Learning

Before AI was used to support pedagogical interventions, SRL was commonly developed through teacher-led strategy training, reflective journals, peer support, and metacognitive prompts. Meta-analyses show consistent benefits: university-level SRL programs improve academic performance and metacognitive strategies (Theobald, 2021), while K-12 interventions yield even larger effects on achievement (Elhusseini et al., 2022). Gains are strongest when instruction emphasizes metacognition and includes feedback from teachers or peers (Jansen et al., 2019).

Both traditional and AI-PI aim to enhance students' self-regulated learning, but they differ in how this support is delivered. Traditional interventions rely on human interaction and the expertise of teachers or mentors to guide learners. This human involvement allows for rich, nuanced feedback and emotional support, such as encouragement and empathy, that technology alone may not fully replicate. In contrast, AI-based systems are well suited for scalability and real-time adaptation. For example, intelligent tutoring systems can continuously track learner performance and immediately tailor practice questions or hints, whereas human instructors face practical limits in providing instant, individualized feedback in classroom settings (Lan & Zhou, 2025).

However, there are also cautions to consider. Some scholars argue that an over-reliance on AI guidance could risk diminishing students' own agency if not carefully managed, for instance, if learners simply follow algorithmic prompts without internalizing the self-regulatory reasoning behind them (Molenaar, 2022). The ideal path forward, as suggested by Molenaar (2022), is to combine the strengths of both approaches. That is "hybrid human-AI regulation" where control over the learning process dynamically shifts between the learner, the human teacher, and the AI as needed, with the aim of transferring responsibility to the learner.

Existing Reviews and Research Questions

Although the use of AI to support SRL is still an emerging area, a substantial body of empirical research has already accumulated. As a result, numerous recent systematic reviews have examined AI-supported SRL from different angles, including reviews with a broad scope (Banihashem et al., 2025) and those focused on specific contexts such as higher education (Lan & Zhou, 2025), language learning (e.g., Chang & Sun, 2024; Zhang et al., 2025), generative AI (Sardi et al., 2025; Qi et al., 2025), AI chatbots (Wu et al., 2025), AI-generated scaffolds (Qian et al., 2025), and data streams and learning analytics (de Mooij et al., 2025). Despite this growing literature, research synthesis that quantifies the effect of AI integration on SRL remains limited. Existing meta-analyses have been predominantly used to evaluate AI tools or interventions on learning outcomes instead of learning strategies (e.g., Dai et al., 2024; Deng et al., 2025; Zheng et al., 2023). To our knowledge, the only meta-analysis directly related to this topic is the one by Yin et al. (2025), which examined the effects of pedagogical agents (a specific type of AI-enhanced learning assistant) on overall SRL. Beyond the work of Yin et al. (2025), our meta-analysis provides a deeper and more comprehensive examination by investigating how different forms of AI tools, such as intelligent tutoring systems, conversational agents, adaptive learning platforms, and learning analytics dashboards, influence the effects of interventions on overall SRL as well as specific SRL subdomains.

The following three research questions guided our investigation.

1. What is the overall effect of AI-based pedagogical interventions (AI-PI) on self-regulated learning (SRL) strategies and how does the effect vary across studies?
2. What are the effects of AI-PI on subdomains (e.g., metacognitive, cognitive, and resource management strategies) of SRL? How do the effects vary across studies?

3. What are the conditions (e.g., educational level, discipline, intervention length, type of AI applications, AI's pedagogical role, SRL measure) that moderate the effect of AI-based interventions on SRL strategies?

Method

Study Identification

Potentially relevant studies were identified through searches in ERIC, Education Source, PsycINFO, Web of Science, Scopus, and ProQuest Dissertations and Theses Global. As documented in Appendix B, our search string included three categories: SRL, AI, and research design. The Boolean operator "and" connected the three categories, and "or" was used within each category. Search terms were developed by reviewing and adapting published search terms for SRL (Broadbent & Poon, 2015; Cheng et al., 2025), AI (Ouyang et al., 2022; Zawacki-Richter et al., 2019), and research design (Deng et al., 2025).

Within the SRL category, we included both general terms for SRL and terms for specific strategies, such as time management and help-seeking, to capture studies that examined individual strategies without explicitly referring to the broader SRL construct. Within the AI category, we expanded the terms to encompass specific models, tools, and applications, accommodating variations in how researchers described AI-based interventions. The research-design block focused retrieval on experimental and quasi-experimental studies, consistent with our aim of estimating intervention effects on SRL. Its inclusion was informed by previous meta-analyses of technology-based interventions (e.g., Deng et al., 2025) that used design terms in their search strategies, as well as by the consideration that experimental designs provide a relatively strong basis for causal inference (e.g., Shadish et al., 2002). To reduce reliance on design terminology appearing in titles or abstracts, this block was searched across broader searchable fields where supported. SRL and AI terms were searched in titles, abstracts, and keywords (where supported by the database) because these central concepts were expected to be reported in those fields.

The search was conducted on December 29, 2024. No initial date limit was applied to ensure a comprehensive inclusion of relevant studies. The complete database-specific search strategies are provided in Appendix B, including field codes, Boolean operators, quotation marks, and truncation symbols. ERIC, Education Source, and PsycINFO were searched simultaneously through the EBSCOhost interface using a common search string. The search terms and Boolean structure were consistent across platforms, while field restrictions were adapted to each platform's search interface. Definitions of the field codes are provided in the appendix notes.

Inclusion and Exclusion Criteria

Inclusion Criteria

(1) Were reported in English
(2) Were reported in journal articles, conference papers, or dissertations.
(3) Employed experimental or quasi-experimental designs and at least one experimental group employed AI-based intervention (e.g., AI-driven feedback, interaction with a conversational agent, adaptive prompts and recommendations)
(4) Studies included a control group that either received no intervention (passive control) or used an alternative traditional pedagogical approach that relied on a teacher, peers, a non-intelligent learning tool or system, textbooks, or checklists (active control)

(5) Included SRL, a SRL subdomain (metacognitive, cognitive, or resource management strategies), or at least one SRL strategy (e.g., help-seeking, self-evaluation) as an outcome variable

(6) Reported sufficient quantitative data (e.g., means and standard deviations, t-statistic, F-statistic) for computing effect size.

Exclusion Criteria

(1) Employed a one-group pre-posttest design, which poses a potential threat to internal validity and limits causal inference

(2) Examined the effect of an AI-PI on multiple SRL strategies but reported quantitative results only for strategies that reached statistical significance. This selective reporting could introduce bias into the evidence base.

Screening

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021) for screening identified reports. The PRISMA flow diagram was used to illustrate the screening process, and Rayyan (Ouzzani et al., 2016), a digital systematic review platform, was used to manage and facilitate screening.

As shown in Figure 1, our search identified 5,741 records, which included 1,323 duplicates. After removing duplicates, 4,418 records remained for the initial round of screening. Given the large number of records, the first and second authors jointly screened an initial set of 50 records to pilot the process and to ensure consistent application of the inclusion and exclusion criteria. After piloting, the two authors divided the remaining records into two sets and independently screened titles, keywords, and abstracts based on the established criteria. Both reviewers adopted a conservative decision rule: records were retained for full-text assessment whenever the available title, keywords, and abstract did not provide sufficient information to determine eligibility.

A total of 216 reports were retained for full-text review. These reports were divided into two groups. The first and fourth authors reviewed reports published in 2024 and 2025, which resulted in 130 records. The second and third authors reviewed reports published in or before 2023, which resulted in 86 records. Each report was reviewed by two authors to ensure reliability. Within each group, the corresponding two authors independently reviewed all full text reports, and any disagreements were resolved through discussion.

After full-text screening, 34 reports met all eligibility criteria and were included for coding and analysis. These reports consisted of 30 journal articles and four conference papers.

Citation Searching

Backward citation searching was conducted using the reference lists of the 34 reports included in the meta-analysis, yielding 2,176 records. Before screening, 251 duplicates or already included records were removed; this count included five reports already included and coded in the meta-analysis. The remaining records were compared with the 4,418 records screened during the initial search, identifying 13 additional previously screened records. In total, 264 records were removed, leaving 1,912 new records for title and abstract screening.

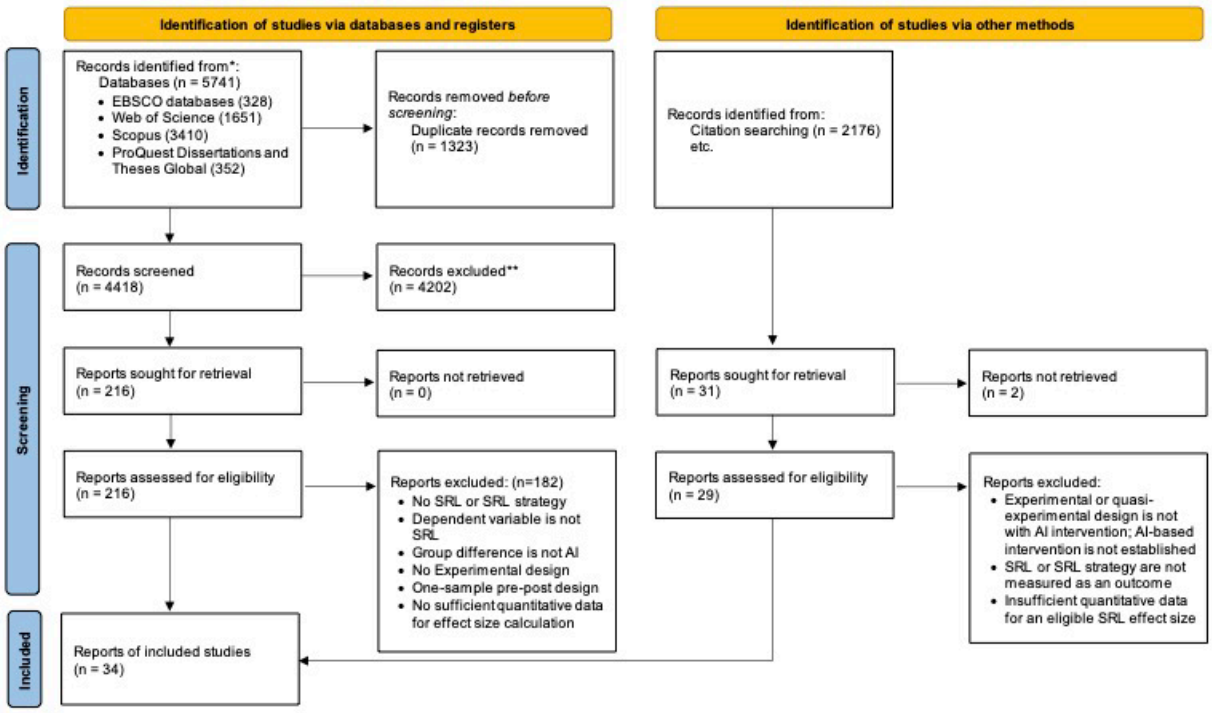
One researcher served as the primary reviewer for title and abstract screening. This researcher had completed screening training and participated in the initial screening round; therefore, no additional pilot screening exercise was conducted. Records were screened against the established inclusion and

exclusion criteria. Of the 1,912 records screened, 1,881 were excluded, and 31 reports were sought for full-text review.

Two dissertations could not be retrieved, leaving 29 reports for full-text assessment by two researchers. The primary reviewer assessed all available reports and provisionally identified two for inclusion. The second researcher served as a co-reviewer, after which the researchers discussed their assessments and reached consensus. Ultimately, all 29 retrieved reports were excluded: 20 did not evaluate an established AI-based intervention using an experimental or quasi-experimental design (e.g., Arnold & Pistilli, 2012); seven did not measure SRL or an SRL strategy as an outcome (e.g., Arguedas et al., 2018); and two provided insufficient quantitative data to calculate an eligible SRL effect size (e.g., Zheng et al., 2019). Thus, backward citation searching yielded no additional reports for inclusion in the meta-analysis.

Figure 1

PRISMA Flow Diagram



Quality Appraisal

The methodological quality of the 34 included articles was appraised using six criteria adapted from Cheng et al. (2025): baseline control, external validity, internal validity, examination of statistical assumptions, construct validity, and power or sample adequacy. The appraisal built on existing study-characteristic codes for pretest administration (pretest versus no pretest) and research design (experimental versus quasi-experimental). Two researchers independently reviewed these codes alongside the full texts and coded the additional information required for the six appraisal criteria. Each criterion was scored dichotomously (1 = criterion met; 0 = criterion not met or not reported), and the ratings were summed to produce a descriptive quality score ranging from 0 to 6.

Before reconciliation, the researchers agreed on 182 of 204 ratings (89.2%). Twenty-two discrepancies occurred across 17 articles, primarily involving baseline control and interpretation of the appraisal criteria. The researchers revisited the relevant full texts, discussed discrepancies using consistent decision rules, and reached consensus on all final ratings. Agreement statistics were calculated from the independent ratings before reconciliation.

The consensus quality scores ranged from 1 to 4 (M = 2.76, SD = 0.85). Three articles scored 1, eight scored 2, seventeen scored 3, and six scored 4.

No additional exclusion threshold was imposed based on quality-appraisal scores. This approach is consistent with examining methodological limitations within the synthesis rather than excluding studies solely based on quality judgments (Glass et al., 1981). Appraisal findings informed the interpretation of the evidence and its limitations.

Coding, Categorization, and Moderators

The codebook consists of five categories of variables: report characteristics (e.g., publication year, report type), research context (e.g., educational level, discipline), research design (e.g., type of experiment, type of control group, intervention length, sample size, pre-test inclusion), AI (e.g., AI applications, pedagogical role), and SRL (e.g., SRL strategies, SRL measure). The 34 retained reports were divided into two parts. The first and second authors coded 28 reports, and the third and fourth authors coded six reports. Each report was coded by two authors to establish interrater reliability. Within each group, the corresponding two authors independently coded all full-text reports. Percentage agreement was calculated separately for each reviewer pair across all coding fields listed in Appendix C. Disagreements between the independently completed coding forms were highlighted and counted, with each discrepant field counted as one disagreement. Percentage agreement was calculated as one minus the number of disagreements divided by the total number of coding fields compared, multiplied by 100. Agreement before discussion was 95.24% for the first and second authors, who coded 28 reports, and 93.33% for the third and fourth authors, who coded six reports. All disagreements were resolved through discussion to produce the final consensus dataset.

The coded textual information was analyzed and categorized below and used for moderator analyses. The codes and their descriptions were also organized into a codebook table in Appendix C.

Research Context Moderators

Three moderators were related to the research context. These were discipline (education, language, engineering, science), whether the study focused on STEM (Science, Technology, Engineering, and Mathematics) or non-STEM, and educational level (elementary, secondary, and higher education).

Research Design Moderators

Five moderators were related to research design. These were type of experiment (experimental or quasi-experimental), intervention length (less than 1 week, 1 to 5 weeks, 6 to 10 weeks, 11 to 16 weeks), sample size (1 to 50, 51 to 100, 101 to 150, and more than 150), whether a pretest was included (yes or no), and type of control group (passive or active control). To guide coding of passive and active control, we first separated studies that provided no instructional support in the control group (passive control) from those that provided some form of instructional support (active control). Passive control referred to studies where the control group did not receive any instructional support from AI, teachers, peers, or other tools. Active control referred to studies in which both groups received support, but the experimental group received AI-based support while the control group received a non-AI alternative.

For example, studies comparing AI-generated feedback with teacher or system feedback were coded as active control because both groups received feedback. However, if only the experimental group received AI-generated feedback and the control group received no feedback, the study was coded as passive control. A more detailed classification of active-control conditions was considered, but subdividing the 17 studies with active controls resulted in sparse categories. We therefore retained the broader active–passive distinction for moderator analysis.

SRL Moderators

Two moderators are associated with SRL strategies. They are SRL phase and SRL measure. Aligned with Zimmerman's three phase cyclic model of SRL (Zimmerman 2000), SRL strategies are categorized into forethought, performance, and reflection phase. In addition, we categorized SRL measures based on when they were administered (Rovers et al., 2019; Schellings 2011; Veenman 2005). Measures collected in real time during a learning task (e.g., trace/log data, observations) were classified as online measures. Measures administered before or after a learning task (e.g., self-report questionnaires) were classified as offline measures.

AI Moderators

Three moderators were associated with AI. These were AI applications, whether the application used Generative AI (yes or no), and the pedagogical role of AI.

For the moderator of AI applications, four categories were used. The first category was conversational agent (CA) (e.g., chatbot, pedagogical agent). These systems interact with learners through natural language and may appear as chatbots, avatars, or named virtual companions such as "Bubi" (Liang et al., 2024) or "Replika" (Ortega-Ochoa et al., 2024). The second category was LLM-supported systems (e.g., ChatGPT, GPT, BERT), which involve the direct use of foundation models as tools, such as for writing or text generation, without a specialized persona (e.g., Chen et al., 2025; Zheng et al., 2021). The third category was intelligent tutoring or adaptive learning systems, which adjust content, difficulty, or pacing based on learner performance using rule-based or machine learning approaches (e.g., Chu et al., 2023). The fourth category was learning analytic systems (e.g., dashboards or visualization tools) that collect and display learning data to support awareness and reflection without directly guiding task completion (e.g., Liao et al., 2024).

For the moderator of the pedagogical role of AI, four roles were identified. The first was feedback or assessment, where AI evaluates student work and provides information on correctness or quality. The second was monitoring or analytics, where AI tracks or visualizes learning behaviors to increase metacognitive awareness without offering evaluative comments (e.g., Liao et al., 2024). The third was scaffolding or facilitation (e.g., prompts, hints, recommendations), where AI guides learning processes by offering strategies or structured support that help students progress in tasks (e.g., Chen et al., 2025). The fourth was mixed, which included systems that combine multiple pedagogical functions and cannot be categorized under a single role (e.g., Liang et al., 2024).

Effect Size

In this meta-analysis, Hedges' g (Hedges, 1981) was used as the measure of the effect of AI-PI on SRL. Hedges' g was calculated as the standardized mean difference between the experimental and control group, with a small-sample bias correction applied. When pre and posttest data were both reported, the effect size was computed using a pretest-adjusted difference approach, which accounts for initial group differences and provides a more accurate estimate of the effect. The pre-posttest correlation was assumed to be .50 when not reported. For studies reporting test statistics (e.g., t, F, or

χ^2) instead of means and standard deviations, Hedges' g was derived from those statistics using established conversion formulas.

Handling Dependence Among Effect Sizes

Because many primary studies (e.g., Fleur et al., 2023; Lim et al., 2023) reported multiple effect sizes for different SRL strategies, these effect sizes from the same study were statistically dependent. Ignoring such dependency can result in underestimated standard errors and inflated Type I error rates (Hedges et al., 2010). To address this issue, the present study employed the robust variance estimation (RVE, Hedges et al., 2010) and implemented in the robumeta package (Fisher & Tipton, 2015). The RVE method extends random-effects meta-regression to accommodate multiple, correlated effect sizes from the same study. It does so by adjusting the weighting and estimating robust standard errors that remain valid even when the exact covariance structure is unknown. A within-study correlation of $\rho = .80$ was assumed among effect sizes (Tanner-Smith & Tipton, 2014), reflecting a moderate to high degree of dependence typically observed in educational and psychological research. The small-sample correction recommended by Tipton (2015) was applied to improve the accuracy of statistical inference given the modest number of studies included in the synthesis.

Data Analysis

Analyses associated with research questions were conducted in R Studio using random-effects models estimated with the RVE method implemented in the robumeta package (Fisher and Tipton, 2015). Random-effects models were used to account for potential variability in true effect sizes across studies, recognizing that the impact of AI-PI on SRL may differ across moderators. As suggested by the RVE method (Hedges et al., 2010), Heterogeneity (variation in the true effect sizes, Borenstein et al., 2021) was measured using the between-study variance (τ^2) and the proportion of variance attributable to heterogeneity (I^2).

The first research question examined the overall effect of AI-PI on SRL as a global construct. A random-effects RVE model was estimated to obtain the Hedges' g across all included studies. The standard error and the 95% confidence interval were reported to indicate the precision of the estimate. Heterogeneity was assessed using τ^2 and I^2 to evaluate whether meaningful variability existed across studies.

The second research question examined whether the effects of AI-PI differed across SRL subdomains (metacognitive, cognitive, and resource management strategies). Three separate random-effects RVE meta-analyses were conducted to estimate the overall Hedges' g for each subdomain.

For the third research question, moderator analyses were also conducted using RVE-based meta-regression models. Each moderator was examined in a separate meta-regression model. In each model, the intercept represented the estimated mean effect size for the reference category, and regression coefficients represented differences between the reference category and other levels of the moderator. Statistical significance of these coefficients was evaluated using t tests to assess whether the effects of AI-PI differed across moderator levels. To obtain all possible pairwise comparisons among moderator levels, the reference category was systematically rotated, and a series of models was estimated for each moderator, with a different level specified as the reference in each model. This procedure yielded a complete set of pairwise coefficient comparisons and allowed for direct examination of differences across all categories. A moderator was considered significant if at least one pair of comparison yielded a statistically significant t test.

Publication bias was examined using the funnel plot, Egger's regression test (Egger et al., 1997), and the trim-and-fill method (Duval & Tweedie, 2000). Analyses were conducted at the study-level to

preserve the assumption of independence underlying funnel plot interpretation. Specifically, multiple effect sizes within studies were aggregated into a single study-level effect size using inverse variance weighting. Publication bias analyses were conducted in R Studio using the metafor package (Viechtbauer, 2010).

Results

A total of 91 independent and dependent effect sizes were retrieved from 34 included studies. Descriptive information about the 34 included studies were shown in Table 2. The 34 studies included in this meta-analysis were published primarily in recent years, with half appearing in 2024 or 2025 (accessible online in 2024) and most of the remainder published between 2021 and 2023. The majority were journal articles, with a small number of conference papers and no dissertations meeting the inclusion criteria. Most studies were conducted in higher education settings, with fewer focusing on secondary or elementary education. The studies covered multiple disciplinary domains, most commonly language learning and science, followed by education and engineering. Both experimental and quasi-experimental designs were common, with slightly more quasi-experimental studies. Active and passive control groups were used at similar rates. Intervention length varied widely, from less than one week to up to 16 weeks, with short-term interventions being most frequent.

Table 2

Descriptive Information for Included Studies (n = 34)

Variables	Numbers
Year of Publication	
2024 & 2025	17
2021 - 2023	10
2020 or before	7
Type of Publication	
Journal article	30
Conference paper	4
Dissertation	0
Educational level	
Higher Education	25
Secondary Education	5
Elementary	4
Discipline	
Engineering	1
Science	9
Education	3
Language	13
Research Design	
Experimental	15
Quasi-experimental	19
Type of Control group	
Active control	17
Passive control	18
Intervention Length (weeks)	
<1	15
1 - 5	4
6 - 10	6
11-16	8

Note 1. Totals do not sum to 34 for some variables because of missing study-level information

Note 2. Totals exceed 34 for control group type because one study included both active and passive controls

RQ1: Overall Effect of AI-Based Interventions on SRL and Heterogeneity

As shown in Table 3, the overall mean effect size of AI-PI on SRL across 34 independent studies (91 effect sizes) was statistically significant (Hedges' g = 0.76, SE = 0.12, p < .001). This indicates a moderate-to-large positive effect (Sawilowsky, 2009) of the AI-PI on SRL.

Between-study variation in effect sizes (heterogeneity) was substantial (I^2 = 88.34%, τ^2 = 0.36), suggesting considerable variability in true effects across studies. The forest plot in Appendix A provides a visual summary of between-study heterogeneity. Given this level of heterogeneity, subsequent moderator analyses were necessary to examine potential study level factors (e.g., educational level, discipline, type of AI application, and AI pedagogical role) that may help explain differences in observed effects.

Table 3

Summary Effect of AI-PI on SRL Overall

Effect Size	Estimate	SE	t(df)	95% CI	p
Overall Hedges' g	0.76	0.12	6.41 (32.7)	[0.52, 1.01]	***

Note. *** p < .001

RQ2 Effects of AI-Based Pedagogical Interventions (AI-PI) on Subdomains of SRL and Heterogeneity

The second research question in this meta-analysis focuses on the effects of AI-PI on the subdomains of SRL including cognitive, metacognitive, and resource management strategies. As shown in Table 4, across the three SRL subdomains, AI-PI demonstrated consistently positive effects, though the magnitude of effects varied.

For metacognitive strategies, the analysis included 20 studies and 38 effect sizes. The pooled effect was large (g = 0.73, p < .001). For cognitive strategies, the model synthesized 9 studies and 15 effect sizes. The pooled effect was moderate (g = 0.48, p < .05). For resource-management strategies, the analysis included 11 studies and 21 effect sizes. The pooled effect was large (g = 0.93, p < .05).

Taken together, the results show that AI-PI had a positive and meaningful impact on all three SRL subdomains. The strongest effects emerged for resource management and metacognitive regulation, while cognitive strategy use showed more moderate improvements.

Table 4

Effects of AI-PI on Subdomains of SRL Strategies

SRL Subdomain	Hedges' g	SE	95% CI	t	df	p
Metacognitive	0.73	0.14	[0.44, 1.02]	5.21	8.00	***
Cognitive	0.48	0.20	[0.04, 0.90]	2.39	7.96	*
Resource Management	0.93	0.30	[0.24, 1.62]	3.02	9.90	*

Note. * p < .05, ** p < .01, *** p < .001

RQ3: Moderator Analyses

Moderators Associated with SRL Strategies

SRL Strategies by Learning Phase. The reflection phase yielded the largest mean effect size (g = 1.02, SE = 0.23), followed by the performance phase (g = 0.86). The forethought phase and the mixed phase both showed moderate effect sizes. Results from the meta-regression models indicated that pairwise t tests comparing regression coefficients across learning phases were not statistically significant ($p > .05$), suggesting that learning phase did not function as a significant moderator of SRL outcomes.

SRL Measure. Although offline measures yielded a larger effect size (g = 0.85) than online measures (g = 0.62), the t-test comparing regression coefficients for online and offline measures were not statistically significant ($p > .05$), indicating that SRL measure did not significantly moderate intervention effects**.**

Table 5

SRL Phase and Measure

Moderator & Subgroups	K (ES)	Mean	SE
SRL Strategies by learning phase			
Forethought Phase	8	0.72	0.33
Performance Phase	45	0.86	0.21
Reflection Phase	9	1.02	0.23
mixed	13	0.73	0.20
SRL Measure			
Online Measure (e.g., log, trace data)	31	0.62	0.12
Offline Measure (e.g., self-report survey)	59	0.85	0.18

Note. SE = standard error, ES = effect size

Moderators Associated with AI

Based on the results from pairwise t tests of regression coefficients across moderator levels, both the type of AI application and the pedagogical role of AI significantly moderated the effects of AI-PI on SRL, whereas the use of generative AI versus non-generative AI did not.

Type of AI application. Intelligent tutoring or adaptive learning systems yielded the largest mean effect size (g = 1.20), followed by conversational agents (g = 0.88), LLM-empowered systems (g = 0.56), and learning analytic systems (g = 0.37). Pairwise comparisons indicated that conversational agent-based interventions were significantly more effective in supporting SRL than learning analytic systems.

Pedagogical role of AI. Feedback or assessment showed the largest mean effect size (g = 1.11), followed by scaffolding or facilitation (g = 0.80), mixed roles (g = 0.71), and monitoring or analytics (g = 0.32). Pairwise comparisons indicated significantly stronger effects for scaffolding or facilitation roles than for monitoring or analytic roles.

Generative AI versus non-generative AI. The t-test comparing regression coefficients for generative AI and non-generative AI were not statistically significant ($p > .05$), indicating using generative AI or not is not a significant moderator.

Table 6

AI Applications and Pedagogical Role of AI

Moderator & Subgroups	K (ES)	Mean	SE
Type of AI			
Conversational Agent (e.g., chatbot, pedagogical agent)	23	0.88	0.15
Intelligent tutoring/adaptive learning system	23	1.20	0.38
Learning analytic system (e.g., learning analytic dashboard)	12	0.37	0.10
LLM-empowered system (e.g., GPT, BERT)	33	0.56	0.19
Pedagogical role of AI			
Feedback/Assessment	18	1.11	0.51
Monitoring/Analytics	6	0.32	0.09
Scaffolding/Facilitation	38	0.80	0.16
Mixed	29	0.71	0.16
Generative AI or not			
Yes	35	0.59	0.14
No	56	0.88	0.17

Note. SE = standard error, ES = effect size

Moderators Associated with Research Context

The moderator analysis showed that educational level functioned as a significant moderator of the effects of AI-PI on SRL, whereas discipline and STEM versus non-STEM affiliation did not.

Educational Level. Effect sizes differed across stages. Higher education showed the largest mean effect size (g = 0.83), followed closely by elementary education (g = 0.81), while secondary education demonstrated a smaller effect (g = 0.41). The elementary-secondary comparison yielded an estimated difference of g = 0.40, which was marginally significant. Higher-secondary and higher-elementary comparisons did not meet the .05 threshold. These findings should be interpreted cautiously, particularly given the borderline elementary-secondary result and the low degrees of freedom for the higher-elementary comparison.

Discipline. The moderator was not significant. Although effect sizes varied descriptively across subject areas, with language learning (g = 0.97), education (g = 0.76), science (g = 0.73), and engineering (g = 0.59), pairwise comparisons did not support significant differences across disciplines.

STEM versus non-STEM. This factor did not significantly moderate the effects of AI-PI on SRL, given that mean effect sizes were comparable between non-STEM disciplines (g = 0.82) and STEM disciplines (g = 0.66).

Table 7

Educational Level and Discipline

Moderator & Subgroups	K (ES)	Mean	SE
Educational Level			
Elementary Education	7	0.81	0.15
Secondary Education	8	0.41	0.08
Higher Education	76	0.83	0.16
Discipline			
Education	5	0.76	0.44
Engineering	6	0.59	0.00
Language	31	0.97	0.28

Moderator & Subgroups	K (ES)	Mean	SE
Science	17	0.73	0.15
STEM or not			
Yes	29	0.66	0.13
No	45	0.82	0.20

Note. SE = standard error, ES = effect size

Moderators Associated with Intervention Design

The moderator analyses examining experimental design, type of control group, pretest inclusion, intervention length, and sample size indicated that none of these study design characteristics functioned as significant moderators of the effects of AI-PI on SRL.

Although these moderators were not statistically significant, descriptive patterns in effect sizes were observed. Effect sizes were comparable for experimental (g = 0.77) and quasi-experimental designs (g = 0.76), as well as for studies using active versus passive control groups. Studies that included pretest measures showed larger mean effect sizes (g = 0.94) than those without pretests (g = 0.59), although this difference was not supported by pairwise comparisons.

For intervention length, shorter interventions (less than one week; g = 0.84) and longer length interventions (11 to 16 weeks; g = 0.85) yielded larger mean effect sizes than moderate length interventions (6 to 10 weeks; g = 0.62), but no pairwise differences were statistically significant. Similarly, sample size did not significantly moderate effects, although larger effect sizes were observed in studies with small to medium samples compared with studies with larger samples.

Table 8

Research Design, Intervention Length, sample size

Moderator & Subgroups	K (ES)	Mean	SE
Experimental or Quasi-experimental			
Experimental	31	.77	.21
Quasi-experimental	60	.76	.14
Type of Control Group			
Active control	44	.73	.16
Passive control	47	.80	.18
Accounted for Pre-test or not			
Yes	40	.94	.21
No	51	.59	.11
Intervention length			
<1week	45	.84	.24
1 – 5 weeks	10	.70	.11
6 – 10 weeks	14	.62	.16
11 – 16 weeks	21	.85	.22
Sample Size			
1 - 50	24	.93	.31
51 - 100	41	.62	.11
101 - 150	14	1.28	.42
> 150	12	.37	.09

Note. SE = standard error, ES = effect size

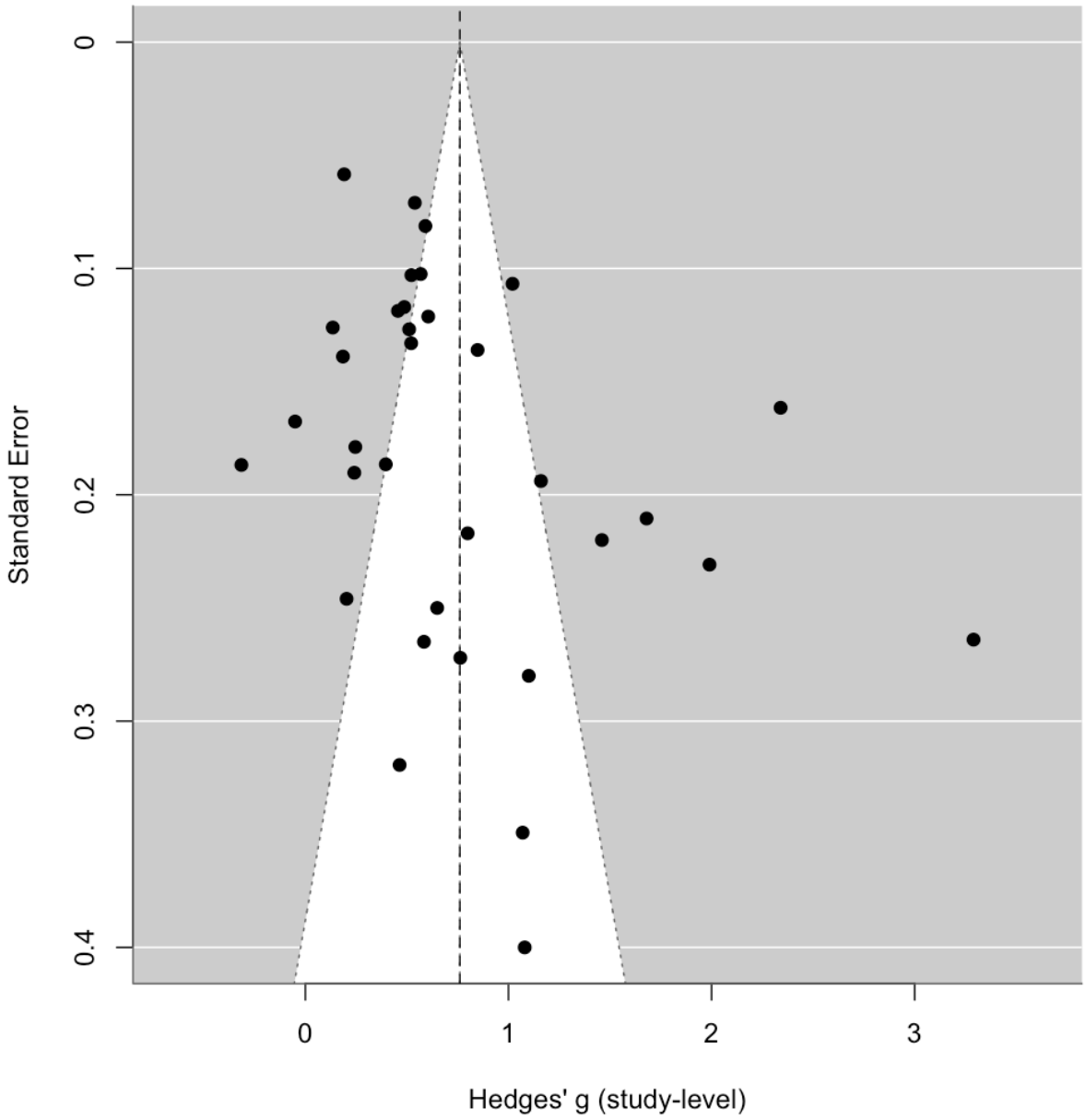
Detailed results for all pairwise moderator comparisons, including estimated differences, robust standard errors, t statistics, degrees of freedom, and unadjusted p-values, are provided in the four tables in Appendix D.

Publication Bias

The funnel plot (shown in Figure 2) suggested some asymmetry, with studies with larger standard errors tending to report larger positive effects. Egger's regression test did not provide statistically significant evidence of funnel plot asymmetry ($p > .05$). Trim and fill analysis did not identify any missing studies and yielded a pooled effect size that was comparable to the original estimate, indicating that potential publication bias was unlikely to substantially influence the overall findings. Taken together, these results suggest that although small study effects cannot be entirely ruled out, there is no strong evidence that publication bias meaningfully threatens the validity of the observed effects.

Figure 2

Funnel Plot



Discussion

This meta-analysis examined the effects of AI-PI on learners' SRL and its subdomains, including cognitive, metacognitive, and resource management strategies. The study also investigated whether these effects varied across four groups of moderators related to research context, intervention design, AI characteristics, and SRL measurement. The discussion interprets the main findings in

relation to prior research, provides implications for practice and research, acknowledges limitations and strengths, and outlines directions for future work.

RQ 1 & 2 Effects of AI-based interventions on SRL and SRL Subdomains

The first part of our meta-analysis showed that AI-PI led to a significant improvement in learners' use of SRL strategies, with a medium to large overall effect (g = .76). This finding is consistent with a previous meta-analysis by Yin et al. (2025), which reported a moderate positive effect of pedagogical agents on SRL, as well as qualitative evidence from recent systematic reviews highlighting the potential of AI for supporting SRL (e.g., Banihashem et al., 2025). At a more detailed level, our analyses of SRL subdomains revealed that AI-PI had significant and positive effects across all three subdomains of SRL (cognitive, metacognitive, and resource management strategies). Although all effects were positive, their magnitudes differed. Resource management showed the strongest effect, followed by metacognitive strategies, while the effect on cognitive strategies was comparatively smaller.

Several factors may help tentatively explain this pattern. First, many AI systems in the included studies were intentionally designed to support learners' resource-management behaviors such as help-seeking (e.g., Chen et al., 2025; Chou et al., 2018). For example, Chen et al. (2025) developed an online learning platform that guided students in diagnosing problems, seeking help, evaluating the quality of that help, and integrating feedback from ChatGPT. Because these designs provide frequent opportunities to practice specific strategies while interacting with AI tools, learners may experience more immediate and consistent benefits in these areas. Second, the ways learners interact with AI systems naturally align with key metacognitive processes such as planning, monitoring, and evaluating their own learning. For instance, tools like learning-analytics dashboards help learners monitor their progress (Liao et al., 2024), and AI-generated prompts can encourage reflection or reconsideration of strategies (e.g., Essel et al., 2024). These forms of support may reinforce the processes within the metacognitive cycle and may explain why metacognitive outcomes responded strongly to AI-PI. Third, cognitive strategies such as rehearsal, elaboration, and organization depend heavily on learners' prior knowledge, personal ways of thinking, and learning beliefs (Pintrich, 2000). These deeper and internal processes often take longer to develop and may not change substantially during short-term interventions. This may help explain the relatively smaller effect observed in the cognitive strategy domain. These explanations, however, remain tentative: they are hypotheses derived from the characteristics of the included interventions and SRL theory, and the present meta-analysis did not directly test these mechanisms.

These results have meaningful implications for educators, instructional designers, and AI system developers. First, integrating AI-PI into instruction may be especially helpful for supporting students' management of both internal resources (e.g., effort, time) and external resources (e.g., learning environments, peers, teachers). Second, higher-order thinking skills such as goal setting, self-monitoring, and self-evaluation appear to benefit more from AI-based support than basic cognitive strategies like rehearsal or elaboration. Third, when the goal is to improve cognitive strategies specifically, AI systems should be intentionally designed to provide direct, content-rich scaffolds for these skills (for example, prompts that encourage elaboration or organization). Pairing AI support with teacher guidance may be another good strategy to increase the likelihood of improving students' use of cognitive strategies.

RQ 3 Moderators

Differences Across Educational Levels

Estimated effects were descriptively largest in higher education (g = 0.83), followed closely by elementary education (g = 0.81), and were smaller in secondary education (g = 0.41). However, only the elementary-secondary comparison met the .05 threshold, and this result was borderline (p = .04996). This result contrasts with findings from another meta-analysis by Yin et al. (2025), which found that educational level was not a significant moderating variable in the relationship between pedagogical agents and SRL. One possible explanation for this discrepancy lies in the scope of AI-PI examined. While Yin et al. focused specifically on pedagogical agents, which typically embedded characters that facilitate learning (Yin et al., 2025; Nguyen, 2023), our study included a broader range of AI-based tools such as adaptive learning systems, personalized learning platforms, conversational agents, and learning analytics dashboard. Differences in the range of AI applications examined may contribute to the differing patterns across reviews, although this explanation was not directly tested.

Furthermore, our dataset shows an uneven distribution of studies across educational levels: 25 studies were conducted in higher education contexts, compared to only four in elementary education and five in secondary education. The uneven distribution of studies limits the precision of comparisons involving elementary and secondary education and constrains the generalizability of the observed pattern. This finding aligns with the conclusions of a systematic review by Banihashem et al. (2025), who observed that AI-SRL research has predominantly focused on higher education, with minimal attention paid to K-12 settings. In their review, 74% of studies were conducted in higher-education contexts, while only 20 studies focused on middle or high school, and just one study addressed primary education. Such an unbalanced research landscape leaves the true effects of AI interventions on SRL in K-12 education uncertain, highlighting the need for more inclusive research across educational levels.

Moderators Associated with AI (AI applications, GenAI or not, AI pedagogical Role)

The exploratory moderator comparisons suggest that specific AI applications and pedagogical roles may be relevant to intervention effectiveness. Differences were detected between conversational agents and learning analytic systems and between scaffolding/facilitation and monitoring/analytics, whereas no statistically significant difference was detected between generative and non-generative AI. However, this pattern does not establish that pedagogical function is more consequential than generative AI status. The findings instead identify specific contrasts that warrant further investigation.

With respect to AI application type, intelligent tutoring and adaptive learning systems had the largest estimated mean effect on SRL, followed by conversational agents such as chatbots and pedagogical agents. Learning analytics dashboards and LLM-supported systems had smaller, though still positive, estimates. A tentative explanation for this descriptive pattern is that these technologies support regulation in different ways. Intelligent tutoring and adaptive systems provide ongoing guidance and adaptive prompts, while conversational agents support SRL through interactive dialogue. In contrast, learning analytics systems primarily present descriptive information that learners must interpret and act upon. The observed difference between conversational agents and learning analytics tools is consistent with the possibility that interactive guidance offers benefits beyond making learning processes visible. This interpretation relates to the work of Yin et al. (2025), whose meta-analysis of pedagogical agents found stronger effects for more interactive and supportive agents. Our findings raise the possibility that a similar pattern may extend to other AI applications. However, this interpretation remains tentative: the only AI application comparison that met the unadjusted .05 threshold was conversational agents versus learning analytic systems. This specific contrast does not establish a general hierarchy of effectiveness across AI applications, and the proposed explanation was not directly tested in the present meta-analysis.

The pedagogical role of AI further clarified these findings. AI interventions that focused on feedback, assessment, or scaffolding had the largest estimated mean effects on SRL, followed by

systems with mixed roles, whereas monitoring or analytics-only tools produced the smallest gains. One tentative explanation for this descriptive pattern is that explicit guidance may help learners translate information about their learning into subsequent actions. Feedback and scaffolding roles directly support core regulatory processes, while monitoring tools place most of the interpretive and planning burden on learners. This interpretation aligns with Yin et al. (2025), who found that pedagogical agents framed as virtual assistants were more effective for SRL than mentor-type or peer-like agents. The scaffolding/facilitation–monitoring/analytics contrast suggests a possible benefit of providing structured guidance beyond displaying learning information. However, the remaining comparisons did not meet the .05 threshold, and these exploratory findings do not establish that feedback, scaffolding, or coaching roles are generally superior.

No statistically significant difference was detected between generative and non-generative AI interventions, although both categories had positive estimated mean effects. This finding does not establish equivalent effectiveness or demonstrate that generative capabilities are unimportant for SRL. Differences in intervention design and the limited evidence base may have constrained our ability to detect differences between these categories.

Generative models may expand the possibilities for supporting SRL by enabling richer dialogue, more flexible explanations, and context-sensitive assistance. Clear pedagogical roles and intentional scaffolding may help translate these capabilities into SRL benefits, although the present meta-analysis did not directly test this possibility. As a tentative design implication, educators and developers could consider how generative AI supports specific instructional functions, such as feedback, scaffolding, or monitoring. Future research should examine whether and how particular generative features interact with these pedagogical roles to support learners' SRL strategies.

Moderators Associated with SRL Phase and Measure

Our moderator analyses indicated that the effects of AI-PI on SRL did not differ significantly across SRL phases or by the type of SRL measure used. With respect to SRL phases, including forethought, performance, reflection, and mixed phases, the absence of significant differences may be partly attributable to the unequal distribution of effect sizes across categories. For example, the performance phase contributed 45 effect sizes, whereas the mixed, forethought, and reflection phases contributed 13, 8, and 9 effect sizes, respectively. Such imbalances can reduce statistical power and limit the ability to detect meaningful differences across SRL phases.

For SRL measures, both online measures (e.g., trace/log data, observations) and offline measures (e.g., self-report questionnaires) showed comparable effects, indicating that measurement type did not significantly influence the impact of AI-PI on SRL. These findings suggest that AI-PI have relatively consistent effects across different SRL phases. While current research often highlights the advantages of log or trace data over self-report questionnaires for assessing SRL (Du et al., 2023), our results indicate that both online and offline measures are appropriate for evaluating the effects of AI-PI on SRL

Moderators Associated with Intervention Design

Unexpectedly, none of the intervention design characteristics examined in this study, including experimental design, sample size, intervention length, pretest inclusion, and type of control group, emerged as significant moderators of the effects of AI based pedagogical interventions on SRL. This finding differs from Yin et al. (2025), who reported that intervention length significantly moderated the effects of pedagogical agents on SRL. One possible explanation for this discrepancy is that Yin et al.'s meta-analysis focused on a single category of AI, whereas the present study synthesized a much broader range of AI tools and systems. It is possible that intervention length plays a more important role for

certain types of AI than for others, and that combining diverse AI applications within a single analysis reduced the ability to detect moderation effects.

Another methodological factor that may have contributed to the absence of significant moderators is the use of robust variance estimation. RVE is well suited for handling dependent effect sizes and provides more accurate standard error estimates by accounting for within study dependence and applying small sample corrections (Tanner-Smith et al., 2016). However, these adjustments also make moderator tests more conservative. As a result, even when some design characteristics were associated with meaningful differences in estimated effects, those differences were less likely to reach statistical significance. This may help explain why these intervention design variables did not emerge as statistically significant moderators in the present study.

From a practical perspective, we did not detect statistically significant pairwise differences by experimental design, intervention length, sample size, pretest inclusion, or control-group type. These findings do not establish equivalent effectiveness or robustness across study designs and implementation conditions. The limited and uneven evidence across categories may have constrained our ability to detect differences.

In particular, the absence of a statistically significant sample-size comparison does not demonstrate that intervention effectiveness is independent of study size or that benefits generalize equally to small- and large-scale implementations. Similarly, nonsignificant comparisons involving pretest inclusion and control-group type do not establish stability across design approaches. Future studies should examine these conditions using rigorous designs and sufficiently large samples.

At the institutional policy level, these findings support considering SRL outcomes alongside academic achievement when evaluating AI-based educational tools. Schools and universities could establish procurement and evaluation criteria that examine how a tool supports learners' planning, monitoring, and reflection, and provide professional development for educators to integrate these functions into instruction.

Strengths, Limitations, and Future Research

A key strength of our meta-analysis is that we examined the effects of AI-PI not only on overall SRL but also on its major subdomains including metacognitive, cognitive, and resource management strategies. Future meta-analyses may consider an even more fine-grained level of analysis by moving from subdomains to specific SRL strategies. For example, researchers could directly evaluate the effects of AI-PI on strategies such as goal setting, help seeking, and time management. Conducting analyses at this level, however, would require a substantially larger pool of studies and effect sizes. In addition, whereas many existing AI-related meta-analyses focus primarily on generative AI (e.g., Zhu et al., 2025) or on a single tool such as ChatGPT (Deng et al., 2025), the present study synthesized evidence across a diverse range of AI systems, including intelligent tutoring systems, conversational agents, LLM-empowered systems, and learning analytics tools. This intentional inclusion provides a more comprehensive picture of how different forms of AI tools could be integrated into instructions to support SRL. Despite the broader field coverage, the research-design block may have reduced sensitivity when eligible studies used different terminology or omitted design terms from searchable fields.

This broad scope also shaped the search and screening process. Diverse SRL and AI terminology and multiple databases were used to reduce the risk of missing relevant studies, yielding 4,418 unique records. However, only 34 studies met the focused eligibility criteria, which required experimental or quasi-experimental evaluations of AI-PI and sufficient quantitative information to calculate effect sizes. Many records did not meet these requirements, and some quantitative reports

lacked usable data. Future reviews could refine search strategies against known eligible studies to improve precision while preserving coverage. The exclusion of reports with insufficient quantitative information remains a limitation.

One limitation of this meta-analysis is the relatively small number of studies that met the inclusion criteria (n = 34), which may limit the generalizability of the findings. We applied small-sample corrections in the robust variance estimation models (Tipton, 2015) and used RVE to retain the available effect sizes (k = 91) while accounting for within-study dependence (Hedges et al., 2010). However, these approaches do not overcome the limited evidence within individual moderator categories. For example, only 17 studies included active controls, and further subdivision by specific comparison condition yielded sparse categories. We therefore retained the broader active-passive classification, although combining heterogeneous active-control conditions limits interpretation of that comparison. Future syntheses with a larger evidence base could examine specific control types. Besides, multiple unadjusted pairwise comparisons increased the potential for Type I error, warranting cautious interpretation.

During the screening stage, following joint calibration of 50 records, title and abstract records were screened by one reviewer each. Although full-text reports were independently assessed by two reviewers, the original decisions were not fully preserved during conflict resolution, preventing calculation of agreement before discussion.

Another limitation of the existing literature on AI-PI for SRL is the lack of clear differentiation among specific AI affordances, such as personalization or adaptation, automation, and real time assistance. Many existing studies (e.g., Chen, 2024; Liu et al., 2024) examine AI interventions at the tool level by comparing learning activities completed with or without AI support. While such designs are useful for establishing whether a given AI tool can support SRL, they provide limited insight into which specific feature or affordance of the tool are responsible for observed effects. As a result, the mechanisms through which AI supports learners' SRL often remain unclear. Future research would benefit from designs that isolate and systematically vary individual AI affordances. For example, a study could compare two versions of the same AI system, one that provides personalized prompts based on learners' performance and another that delivers the same prompts in a non-personalized manner. Similarly, researchers could manipulate the timing of AI-support by comparing real-time assistance with delayed feedback while holding other features constant. Such designs would allow researchers to more directly examine how a specific AI affordance contribute to the development of SRL.

Another direction for future research involves expanding the focus of meta-analyses in the field of AI in education. Most existing meta-analyses examine how AI-PI influence academic achievement (e.g., Dong et al., 2025; Zheng, et al., 2023). However, meta-analyses evaluating the impact of AI on learners' behaviors, beliefs, or motivations are still very limited, which are areas that are central to understanding how students engage with and benefit from AI-supported learning environments. More meta-analyses in these areas are needed.

Conclusion

In this paper, we report a meta-analysis of 34 empirical studies on AI-PI's effects on SRL, exploring how these effects vary across studies, subdomains, and moderating variables. We found that, overall, AI-PI showed positive effects on SRL and its subdomains, including cognitive, metacognitive, and resource management strategies. In addition, the effectiveness of these interventions depended less on the presence of AI itself and more on how AI was designed and used instructionally. Differences in effects were most strongly associated with educational level, the pedagogical roles assumed by AI systems, and the types of AI applications implemented. In contrast, factors such as discipline, intervention length, study design characteristics, and measurement approaches did not consistently

influence outcomes. Together, these findings suggest that AI can support SRL across diverse instructional contexts while underscoring the central role of pedagogical design in shaping its impact.

A key contribution of this study is in understanding strategies for integrating AI into education in improving SRL, and in what pedagogical ways AI can best support the different phases and domains of SRL. With the widespread application of AI tools in education, many researchers have argued there is a critical need for improved teacher development in how to use these tools effectively (e.g., Pham & Sampson, 2022). However, much of the discussion on teacher training about GenAI has focused on what these tools can do, and how to use them for cognitive support. Our study found that there is substantial value also in AI providing SRL support, and we recommend that training for educators and students focus more on the why and how of AI by showing how these tools can be used effectively to support people in owning their own learning through self-regulation and metacognition. Doing so can help education move from an AI-directed paradigm towards more AI-empowered learning (Ouyang & Jiao, 2021).

Conflict of Interest

The authors declare that there are no conflicts of interest.

AI Use Statement

The authors used ChatGPT to improve the readability and language of some sentences within the manuscript. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content.

Data Availability

The data are available from the corresponding author upon reasonable request.

References

Studies included in the meta-analysis are marked with an asterisk*

Alwaqdani, M. (2025). Investigating teachers' perceptions of artificial intelligence tools in education: potential and difficulties. Education and Information Technologies, 30(3), 2737-2755. https://doi.org/10.1007/s10639-024-12903-9

*Al-Abdullatif, A. M., Al-Dokhny, A. A., & Drwish, A. M. (2023). Implementing the Bashayer chatbot in Saudi higher education: measuring the influence on students' motivation and learning strategies. *Frontiers in psychology, 14*, 1129070. https://doi.org/10.3389/fpsyg.2023.1129070

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). Association for Computing Machinery. https://doi.org/10.1145/2330601.2330666

Arguedas, M., Xhafa, F., Casillas, L., Daradoumis, T., Peña, A., & Caballé, S. (2018). A model for providing emotion awareness and feedback using fuzzy logic in online learning. *Soft Computing, 22*(3), 963–977. https://doi.org/10.1007/s00500-016-2399-0

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), 50. https://doi.org/10.1186/s41239-025-00548-8

Bauer, E., Greisel, M., Kuznetsov, I., Berndt, M., Kollar, I., Dresel, M., ... & Fischer, F. (2023). Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda. *British Journal of Educational Technology*, *54*(5), 1222-1245. https://doi.org/10.1111/bjet.13336

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis.* John Wiley & Sons.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & 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

Chang, W. L., & Sun, J. C. Y. (2024). Evaluating AI's impact on self-regulated language learning: A systematic review. *System, 126*, 103484. https://doi.org/10.1016/j.system.2024.103484

*Chen, A., Xiang, M., Zhou, J., Jia, J., Shang, J., Li, X., ... & Fan, Y. (2025). Unpacking help-seeking process through multimodal learning analytics: A comparative study of ChatGPT vs Human expert. *Computers & Education, 226*, 105198. https://doi.org/10.1016/j.compedu.2024.105198

*Chen, M. R. A. (2024). The AI chatbot interaction for semantic learning: A collaborative note-taking approach with EFL students. *Language Learning & Technology, 28*(1), 1-25. https://hdl.handle.net/10125/73586

Cheng, Z., Zhang, Z., Xu, Q., Maeda, Y., & Gu, P. (2025). A meta-analysis addressing the relationship between self-regulated learning strategies and academic performance in online higher education. *Journal of Computing in Higher Education*, *37*(1), 195-224. https://doi.org/10.1007/s12528-023-09390-1

*Chu, S. T., Hwang, G. J., Chien, S. Y., & Chang, S. C. (2023). Incorporating teacher intelligence into digital games: An expert system-guided self-regulated learning approach to promoting EFL students' performance in digital gaming contexts. *British Journal of Educational Technology, 54*(2), 534-553. https://doi.org/10.1111/bjet.13260

*Chou, C. Y., Lai, K. R., Chao, P. Y., Tseng, S. F., & Liao, T. Y. (2018). A negotiation-based adaptive learning system for regulating help-seeking behaviors. *Computers & Education, 126,* 115-128. https://doi.org/10.1016/j.compedu.2018.07.010

Dai, C. P., Ke, F., Pan, Y., Moon, J., & Liu, Z. (2024). Effects of artificial intelligence-powered virtual agents on learning outcomes in computer-based simulations: A meta-analysis. *Educational Psychology Review, 36*(1), 31. https://doi.org/10.1007/s10648-024-09855-4

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational psychology review, 28*(3), 425-474. https://doi.org/10.1007/s10648-015-9320-8

Dong, L., Tang, X., & Wang, X. (2025). Examining the effect of artificial intelligence in relation to students' academic achievement in classroom: A meta-analysis. *Computers and Education: Artificial Intelligence*, 100400. https://doi.org/10.1016/j.caeai.2025.100400

Du, J., Hew, K. F., & Liu, L. (2023). What can online traces tell us about students' self-regulated learning? A systematic review of online trace data analysis. *Computers & Education, 201*, 104828. https://doi.org/10.1016/j.compedu.2023.104828

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905

*Dever, D. A., Sonnenfeld, N. A., Wiedbusch, M. D., & Azevedo, R. (2022). Pedagogical agent support and its relationship to learners' self-regulated learning strategy use with an intelligent tutoring system. In *International conference on artificial intelligence in education* (pp. 332-343). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_27

de Mooij, S., Lämsä, J., Lim, L., Aksela, O., Athavale, S., Bistolfi, I., ... & Molenaar, I. (2025). A Systematic Review of Self-Regulated Learning through Integration of Multimodal Data and Artificial Intelligence. *Educational Psychology Review, 37*(2), 54. https://doi.org/10.1007/s10648-025-10028-0

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Elom, C. O., Ayanwale, M. A., Ukeje, I. O., Offiah, G. A., Umoke, C. C., & Ogbonnaya, C. E. (2025). Does AI Knowledge Encourage Cheating? Investigating Student Perceptions, Ethical Engagement, and Academic Integrity in the Digital Age. *International Journal of Learning, Teaching and Educational Research*, *24*(4), 708-729. https://doi.org/10.26803/ijlter.24.4.33

*Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based

conversational large language models (LLMs). *Computers and Education: Artificial Intelligence, 6,* 100198. https://doi.org/10.1016/j.caeai.2023.100198

Elhusseini, S. A., Tischner, C. M., Aspiranti, K. B., & Fedewa, A. L. (2022). A quantitative review of the effects of self-regulation interventions on primary and secondary student academic achievement. Metacognition and Learning, 17(3), 1117–1139. https://doi.org/10.1007/s11409-022-09311-0

*Fleur, D. S., Marshall, M., Pieters, M., Brouwer, N., Oomens, G., Konstantinidis, A., ... & van Vliet, E. A. (2023). IguideME: Supporting Self-Regulated Learning and Academic Achievement with Personalized Peer-Comparison Feedback in Higher Education. *Journal of Learning Analytics, 10*(2), 100-114. https://doi.org/10.18608/jla.2023.7853

Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. arXiv preprint arXiv:1503.02220. https://doi.org/10.48550/arXiv.1503.02220

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*, 4493–4518. https://doi.org/10.1007/s10639-024-12881-y

Gunbatar, S. A., Sirin, G. T., Ilkyaz, O. C., & Mutlu, Y. (2025). Exploring the artificial intelligence interaction profiles of participants with different levels of teaching experience for lesson planning in the context of acids and bases. *Chemistry Education Research and Practice*, *26*(4), 977-995. https://doi.org/10.1039/D5RP00064E

*Gombert, S., Fink, A., Giorgashvili, T., Jivet, I., Di Mitri, D., Yau, J., ... & Drachsler, H. (2024). From the automated assessment of student essay content to highly informative feedback: A case study. *International Journal of Artificial Intelligence in Education, 34*(4), 1378-1416. https://doi.org/10.1007/s40593-023-00387-6

*Giorgashvili, T., Jivet, I., Artelt, C., Biedermann, D., Bengs, D., Goldhammer, F., ... & Drachsler, H. (2024). Exploring Learners' Self-reflection and Intended Actions After Consulting Learning Analytics Dashboards in an Authentic Learning Setting. In: Ferreira Mello, R., Rummel, N., Jivet, I., Pishtari, G., Ruipérez Valiente, J.A. (eds) *Technology Enhanced Learning for Inclusive and Equitable Quality Education*. EC-TEL 2024. Lecture Notes in Computer Science, vol 15159. Springer, Cham. https://doi.org/10.1007/978-3-031-72315-5_10

Glass, G. V., McGaw, B., Smith, M. L. (1981). *Meta-Analysis in Social Research*. United Kingdom: SAGE Publications.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods, 1*(1), 39-65. https://doi.org/10.1002/jrsm.5

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hemmler, Y. M., & Ifenthaler, D. (2024). Self-regulated learning strategies in continuing education: A systematic review and meta-analysis. *Educational Research Review, 45*, 100629. https://doi.org/10.1016/j.edurev.2024.100629

Jansen, R. S., van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. *Educational Research Review, 28*, 100292. https://doi.org/10.1016/j.edurev.2019.100292

*Karaoğlan Yılmaz, F. G., Olpak, Y. Z., & Yılmaz, R. (2018). The effect of the metacognitive support via pedagogical agent on self-regulation skills. *Journal of Educational Computing Research, 56*(2), 159-180. https://doi.org/10.1177/0735633117707696

*Kautzmann, T. R., & Jaques, P. A. (2019). Effects of adaptive training on metacognitive knowledge monitoring ability in computer-based learning. *Computers & Education, 129*, 92-105. https://doi.org/10.1016/j.compedu.2018.10.017

Kim, N. J., & Kim, M. K. (2022). Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing. *Frontiers in education, 7*, 755914. Frontiers Media SA. https://doi.org/10.3389/feduc.2022.755914

Konstantivo, L. V., Vorozhikhin, V. V., Petrov, A. M., Titova, E. S., & Shtykhno, D. A. (2023). Generative artificial intelligence in education: discussions and forecasts. *Open education*, *27*(2), 36-48. https://doi.org/10.21686/1818-4243-2023-2-36-48

Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *Science of Learning, 10*(1), 21. https://doi.org/10.1038/s41539-025-00319-0

*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

*Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., ... & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior, 139*, 107547. https://doi.org/10.1016/j.chb.2022.107547

*Lim, L., Bannert, M., van der Graaf, J., Fan, Y., Rakovic, M., Singh, S., ... & Gašević, D. (2024). How do students learn with real-time personalized scaffolds?. *British Journal of Educational Technology, 55*(4), 1309-1327. https://doi.org/10.1111/bjet.13414

*Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D., & Ye, X. D. (2024). Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 1-25. https://doi.org/10.1080/09588221.2024.2389923

*Liang, H. Y., Hwang, G. J., Hsu, T. Y., & Yeh, J. Y. (2024). Effect of an AI-based chatbot on students' learning performance in alternate reality game-based museum learning. *British Journal of Educational Technology, 55*(5), 2315-2338. https://doi.org/10.1111/bjet.13448

*Long, Y., & Aleven, V. (2013). Active learners: Redesigning an intelligent tutoring system to support self-regulated learning. In *European Conference on Technology Enhanced Learning* (pp. 490-495). Berlin, Heidelberg: Springer Berlin Heidelberg.

Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence, 3*, 100070. https://doi.org/10.1016/j.caeai.2022.100070

*McCarthy, K. S., Likens, A. D., Johnson, A. M., Guerrero, T. A., & McNamara, D. S. (2018). Metacognitive overload!: Positive and negative effects of metacognitive prompts in an intelligent tutoring system. *International Journal of Artificial Intelligence in Education, 28*(3), 420-438. https://doi.org/10.1007/s40593-018-0164-5

Nedungadi, P., Tang, K. Y., & Raman, R. (2024). The transformative power of generative artificial intelligence for achieving the sustainable development goal of quality education. *Sustainability*, *16*(22), 9779. https://doi.org/10.3390/su16229779

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology, 55*(4), 1328-1353. https://doi.org/10.1111/bjet.13454

*Ortega-Ochoa, E., Pérez, J. Q., Arguedas, M., Daradoumis, T., & Puig, J. M. 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,* 101101. https://doi.org/10.1016/j.iot.2024.101101

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011–2020. *Education and Information Technologies, 27*, 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic reviews*, *5*(1), 210.

https://doi. org/10.1186/s13643-016-0384-4

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*, 100020. https://doi.org/10.1016/j.caeai.2021.100020

* Pan, M., Lai, C., & Guo, K. (2025). Effects of GenAI-empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading. *The Internet and Higher Education, 65*, 100991. https://doi.org/10.1016/j.iheduc.2024.100991

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., Hrobjartsson, ´ 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*, n71. https://doi.org/10.1136/bmj.n71

Pham, S. T., & Sampson, P. M. (2022). The development of artificial intelligence in education: A review in context. *Journal of Computer Assisted Learning, 38*(5), 1408–1421. https://doi.org/10.1111/jcal.12687

Pintrich, P. R. (2004). Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students. *Educational Psychology Review 16,* 385-407. https://doi.org/10.1007/s10648-004-0006-x

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International journal of educational research*, *31*(6), 459-470. https://doi.org/10.1016/S0883-0355(99)00015-4

Qian, K., Liu, S., Li, T., Raković, M., Li, X., Guan, R., ... & Gašević, D. (2025). Towards reliable generative AI-driven scaffolding: Reducing hallucinations and enhancing quality in self-regulated learning support. *Computers & Education, 240,* 105448. https://doi.org/10.1016/j.compedu.2025.105448

Qi, X. I. A., Liu, Q., Tlili, A., & Thomas, K. F. (2025). A systematic literature review on designing self-regulated learning using generative artificial intelligence and its future research directions. *Computers & Education, 240*, 105465. https://doi.org/10.1016/j.compedu.2025.105465

*Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology, 14*, 1255594. https://doi.org/10.3389/fpsyg.2023.1255594

Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merri¨enboer, J. J. G. (2019). Granularity matters: Comparing different ways of measuring self-regulated learning. *Metacognition and Learning, 14*(1), 1–19. https://doi.org/10.1007/s11409-019-09188-6

*Ramachandran, A., Litoiu, A., & Scassellati, B. (2016). Shaping productive help-seeking behavior during robot-child tutoring interactions. In *Proceedings of the 11th ACM/IEEE international Conference on human robot interaction* (pp. 247–254). Piscataway, NJ: IEEE Press.

Sardi, J., Darmansyah, Candra, O., Yuliana, D. F., Habibullah, Yanto, D. T. P., & Eliza, F. (2025). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *International Journal of Engineering Pedagogy (iJEP), 15*(1), 94–108. https://doi.org/10.3991/ijep.v15i1.53379

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of modern applied statistical methods, 8*(2), 597–599. https://doi.org/10.22237/jmasm/1257035100

Schellings, G. L. M. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning, 6*(2), 91–109. https://doi.org/10.1007/s11409-011-9069-5

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton, Mifflin and Company

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence*, *4*, 100124. https://doi.org/10.1016/j.caeai.2023.100124

*Serrano, M. Á., Vidal-Abarca, E., & Ferrer, A. (2018). Teaching self-regulation strategies via an intelligent tutoring system (TuinLECweb): Effects for low-skilled comprehenders. Journal of Computer Assisted Learning, 34(5), 515-525. https://doi.org/10.1111/jcal.12256

*Shi, L. (2025). The integration of advanced AI-enabled emotion detection and adaptive learning systems for improved emotional regulation. Journal of Educational Computing Research, 63(1), 173-201. https://doi.org/10.1177/07356331241296890

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, 2(1), 85-112. https://doi.org/10.1007/s40865-016-0026-5

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research synthesis methods, 5*(1), 13-30. https://doi.org/10.1002/jrsm.1091

*Tang, L., Shen, K., Le, H., Shen, Y., Tan, S., Zhao, Y., ... & Fan, Y. (2024). Facilitating learners' self-assessment during formative writing tasks using writing analytics toolkit. *Journal of Computer Assisted Learning, 40*(6), 2822-2839. https://doi.org/10.1111/jcal.13036

*Teng, M. F. (2025). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education, 60*(1), e12811. https://doi.org/10.1111/ejed.12811

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological methods, 20*(3), 375. https://doi.org/10.1037/met0000011

Thottoli, M. M., Alruqaishi, B. H., & Soosaimanickam, A. (2024). Robo academic advisor: Can chatbots and artificial intelligence replace human interaction?. *Contemporary Educational Technology*, *16*(1), ep485. https://doi.org/10.30935/cedtech/13948

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 e101976. https://doi.org/10.1016/j.cedpsych.2021.101976

*Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. Computers & Education, 215, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 77–99). Münster: Waxmann

Viechtbauer, W. (2010). Conducting meta-analyses in R with the met-afor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

*Wang, H. H., & Wang, C. H. A. (2024). Teaching design students machine learning to enhance motivation for learning computational thinking skills. Acta Psychologica, 251, 104619. https://doi.org/10.1016/j.actpsy.2024.104619

Warr, M., Oster, N. J., & Isaac, R. (2024). Implicit bias in large language models: Experimental proof and implications for education. *Journal of research on technology in education*, 1-24. https://doi.org/10.1080/15391523.2024.2395295

*Wei, L. (2023). Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. Frontiers in Psychology, 14, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Wiley, D. (2025). Democratizing AI participation in AI in education. https://opencontent.org/blog/archives/7821

Wu, X. Y., Radloff, J. D., Yeter, I. H., Wang, L., & Chiu, T. K. (2025). Designing artificial intelligence chatbots for self-regulated learning from a systematic review based on Habermas's three interests. Interactive Learning Environments, 1-24. https://doi.org/10.1080/10494820.2025.2563086

*Wu, T. T., Lee, H. Y., Li, P. H., Huang, C. N., & Huang, Y. M. (2024). Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid. Journal of Educational Computing Research, 61(8), 1539-1567. https://doi.org/10.1177/07356331231191125.

Weinstein, C. E., Husman, J., & Dierking, D. R. (2000). Self-Regulation interventions with a focus on learning strategies. In M. Boekaerts, P. R. Printrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 727–747). San Diego, CA, US: Academic Press. https://doi.org/10.1016/b978-012109890-2/50051-2

West, R. E., & Wiley, D. (2024). The 5Cs of openness: Articulating an open education infrastructure (OER). In R. E. West & H. Leary (Eds.), Foundations of learning and instructional design technology: Historical roots & current trends. EdTech Books. https://doi.org/10.59668/473

Yin, X., Ye, J., Yu, S., Luo, S., Nan, M., Zhao, G., ... & Zheng, L. (2025). A Meta-analysis of Pedagogical Agents promoting Self-Regulated Learning. In 2025 7th International Conference on Computer Science and Technologies in Education (CSTE) (pp. 413-417). IEEE. https://doi.org/10.1109/CSTE64638.2025.11091971

*Yıldız, T. A. (2023). The impact of ChatGPT on language learners' motivation. Journal of Teacher Education and Lifelong Learning, 5(2), 582-597. https://doi.org/10.51535/tell.1314355

Yu, H., & Guo, Y. (2023, June). Generative artificial intelligence empowers educational reform: current status, issues, and prospects. In *Frontiers in Education* (Vol. 8, p. 1183162). Frontiers Media SA. https://doi.org/10.3389/feduc.2023.1183162

Yue, M., Jong, M. S. Y., & Ng, D. T. K. (2024). Understanding K–12 teachers' technological pedagogical content knowledge readiness and attitudes toward artificial intelligence education. *Education and information technologies*, *29*(15), 19505-19536. https://doi.org/10.1007/s10639-024-12621-2

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), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. *In Handbook of self-regulation* (pp. 13-39). Academic press.

Zhang, Z., Zhou, S., & Bai, B. (2025). A Systematic Scoping Review of Self-Regulated Learning With AI in Language Education. *International Journal of Applied Linguistics*. 1-15. https://doi.org/10.1111/ijal.70030

Zheng, L., Niu, J., Zhong, L., & Gyasi, J. F. (2023). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments, 31*(9), 5650-5664. https://doi.org/10.1080/10494820.2021.2015693

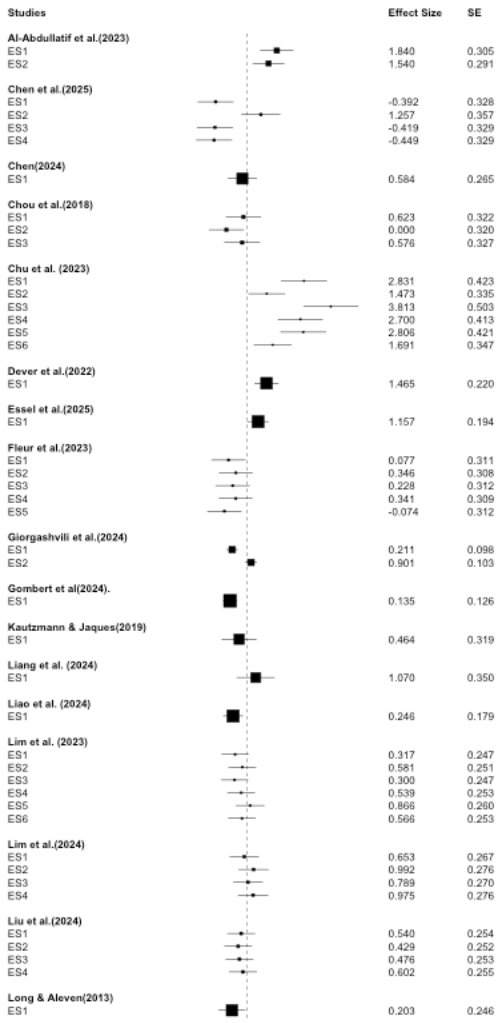
*Zheng, L., Zhong, L., Niu, J., Long, M., & Zhao, J. (2021). Effects of personalized intervention on collaborative knowledge building, group performance, socially shared metacognitive regulation, and cognitive load in computer-supported collaborative learning. *Educational Technology & Society, 24*(3), 174-193. https://www.jstor.org/stable/27032864

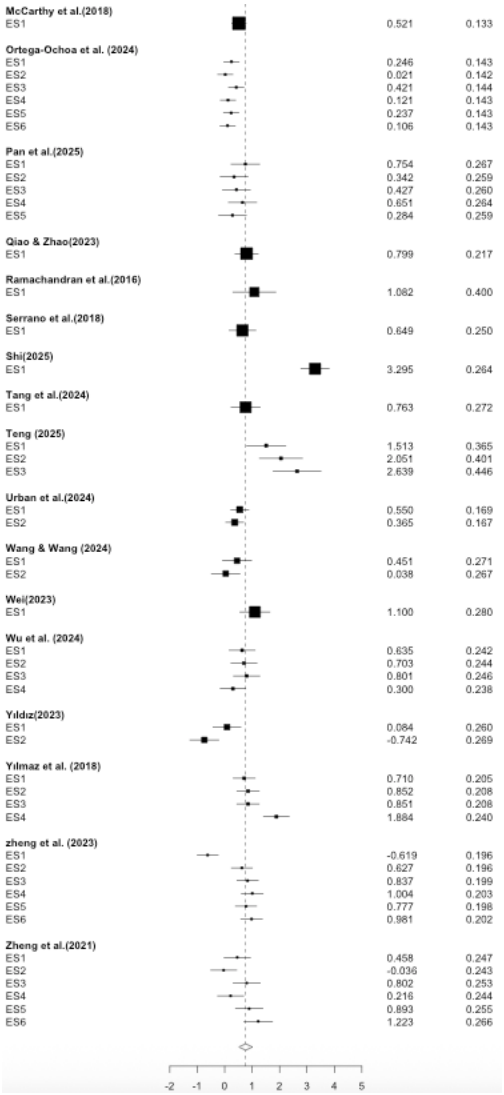
*Zheng, L., Long, M., Chen, B., & Fan, Y. (2023). Promoting knowledge elaboration, socially shared regulation, and group performance in collaborative learning: An automated assessment and feedback approach based on knowledge graphs. *International Journal of Educational Technology in Higher Education, 20*(1), 46. https://doi.org/10.1186/s41239-023-00415-4

Zheng, L., Li, X., Zhang, X., & Sun, W. (2019). The effects of group metacognitive scaffolding on group metacognitive behaviors, group performance, and cognitive load in computer-supported collaborative learning. *The Internet and Higher Education, 42*, 13–24. https://doi.org/10.1016/j.iheduc.2019.03.002

Zhu, Y., Liu, Q., & Zhao, L. (2025). Exploring the impact of generative artificial intelligence on students' learning outcomes: A meta-analysis. *Education and Information Technologies*, 1-29. https://doi.org/10.1007/s10639-025-13420-z

Appendix A. (Forest Plot of Studies 1-17 (Left) and Studies 18-34 (Right))





Appendix B. Search Strings for Different Databases

Databases	Search String
ERIC, Education Source, PsycINFO (EBSCOhost)	AB (“self regulat*” OR “metacog*” OR “learning regul*” OR “goal setting” OR “self evaluat*” OR “self assess*” OR “self reflect*” OR “rehearsal” OR “critical thinking” OR “elaboration” OR “effort regulat*” OR “time management” OR “peer learning” OR “help seeking” OR “environment structur*”) AND AB (“artificial intelligence” OR “AI” OR “large language model” OR “LLM” OR “machine learning” OR “deep learning” OR “expert system” OR “natural language processing” OR “NLP” OR “chatgpt” OR “intelligent agent” OR “intelligent assistant” OR “intelligent tutor” OR “intelligent tutoring system” OR “real time analytic*” OR “adaptive learning system*” OR “automatic feedback” OR “chatbot*” OR “conversational agent”) AND (“experiment*” OR “quasi experiment*” OR “intervention” OR “randomized controlled trial*” OR “randomised controlled trial*” OR "RCT" OR “control group” OR “pre test” OR “post test”)
Scopus	TITLE-ABS-KEY ("self regulat*" OR "metacog*" OR "learning regul*" OR "goal setting" OR "self evaluat*" OR "self assess*" OR "self reflect*" OR "rehearsal" OR "critical thinking" OR "elaboration" OR "effort regulat*" OR "time management" OR "peer learning" OR "help seeking" OR "environment structur*") AND TITLE-ABS-KEY ("artificial intelligence" OR "AI" OR "large language model" OR "LLM" OR "machine learning" OR "deep learning" OR "expert system" OR "natural language processing" OR "NLP" OR "chatgpt" OR "intelligent agent" OR "intelligent assistant" OR "intelligent tutor" OR "intelligent tutoring system" OR "real time analytic*" OR "adaptive learning system*" OR "automatic feedback" OR "chatbot*" OR "conversational agent") AND ALL ("experiment*" OR "quasi experiment*" OR "intervention" OR "randomized controlled trial*" OR "randomised controlled trial*" OR "RCT" OR "control group" OR "pre test" OR "post test")
Web of Science	((TS=(“self regulat*” OR “metacog*” OR “learning regul*” OR “goal setting” OR “self evaluat*” OR “self assess*” OR “self reflect*” OR “rehearsal” OR “critical thinking” OR “elaboration” OR “effort regulat*” OR “time management” OR “peer learning” OR “help seeking” OR “environment structur*”)) AND TS=(“artificial intelligence” OR “AI” OR “large language model” OR “LLM” OR “machine learning” OR “deep learning” OR “expert system” OR “natural language processing” OR “NLP” OR “chatgpt” OR “intelligent agent” OR

Databases	Search String
	“intelligent assistant” OR “intelligent tutor” OR “intelligent tutoring system” OR “real time analytic*” OR “adaptive learning system*” OR “automatic feedback” OR “chatbot*” OR “conversational agent”)) AND TS=(“experiment*” OR “quasi experiment*” OR “intervention” OR “randomized controlled trial*” OR “randomised controlled trial*” OR "RCT" OR “control group” OR “pre test” OR “post test”)
ProQuest Dissertations and Theses Global	noft(“self regulat*” OR “metacog*” OR “learning regul*” OR “goal setting” OR “self evaluat*” OR “self assess*” OR “self reflect*” OR “rehearsal” OR “critical thinking” OR “elaboration” OR “effort regulat*” OR “time management” OR “peer learning” OR “help seeking” OR “environment structur*”) AND noft(“artificial intelligence” OR “AI” OR “large language model” OR “LLM” OR “machine learning” OR “deep learning” OR “expert system” OR “natural language processing” OR “NLP” OR “chatgpt” OR “intelligent agent” OR “intelligent assistant” OR “intelligent tutor” OR “intelligent tutoring system” OR “real time analytic*” OR “adaptive learning system*” OR “automatic feedback” OR “chatbot*” OR “conversational agent”) AND (“experiment*” OR “quasi experiment*” OR “intervention” OR “randomized controlled trial*” OR “randomised controlled trial*” OR "RCT" OR “control group” OR “pre test” OR “post test”)

Note. AB = Abstract in EBSCOhost. TITLE-ABS-KEY = Title, Abstract, and Keywords in Scopus; ALL searches across Scopus’s searchable record fields and does not denote article full-text searching. TS = Topic in Web of Science covering titles, abstracts, and Indexing. NOFT = Anywhere except full text in ProQuest.

Appendix C. Codebook

Category	Variable	Coding options
Report characteristics	Publication year	Year of publication
Report characteristics	Report type	Journal articles, conference papers, dissertations, etc.
Research context	Discipline	Education; language; engineering; science
Research context	STEM affiliation	STEM; non-STEM
Research context	Educational level	Elementary; secondary; higher education
Research design	Type of experiment	Experimental; quasi-experimental
Research design	Intervention length	Less than 1 week; 1–5 weeks; 6–10 weeks; 11–16 weeks
Research design	Sample size	1–50; 51–100; 101–150; more than 150
Research design	Pretest inclusion	Yes; no
Research design	Type of control group	Passive; active
SRL	SRL phase	Forethought; performance; reflection; mixed
SRL	SRL measure	Online; offline
AI	AI application	Conversational agent; LLM-supported system; intelligent tutoring or adaptive learning system; learning analytic system
AI	Generative AI use	Yes; no
AI	Pedagogical role of AI	Feedback or assessment; monitoring or analytics; scaffolding or facilitation; mixed

Appendix D

Table D1

Research context

Comparison A − B	Difference in g	Robust SE	t	df	p
Educational level (34 studies; 91 effect sizes)					
Higher − elementary	0.0245	0.214	0.114	3.92*	.9145168
Elementary − secondary	0.4002	0.1657	2.415	6.35	.04996
Higher − secondary	0.4248	0.1763	2.410	5.75	.054360
Discipline (26 studies; 59 effect sizes)					
Engineering − education	−0.1639	0.441	−0.3713	1.99*	.746
Language − education	0.2159	0.522	0.4138	3.07*	.70623
Science − education	−0.0266	0.467	−0.0568	3.56*	.95772
Engineering − language	−0.380	0.278	−1.366	11.95	.19716
Science − language	−0.2424	0.317	−0.7638	17.29	.45530
Engineering − science	−0.1373	0.153	−0.8973	7.96	.39588
STEM affiliation (30 studies; 74 effect sizes)					
STEM − non-STEM	−0.161	0.242	−0.667	20.84	.512103

Note. Differences are A minus B; positive values indicate a larger estimated effect for category A. SE = standard error; df = small-sample-adjusted degrees of freedom. Duplicate reverse contrasts are omitted.

* df < 4: inference may be unreliable, as flagged in the R output.

Table D2

Research design

Comparison A − B	Difference in g	Robust SE	t	df	p
Study design (34 studies; 91 effect sizes)					
Quasi-experimental − experimental	−0.0161	0.251	−0.0641	30.1	.9493359
Intervention length (33 studies; 90 effect sizes)					
1–5 weeks − 11–16 weeks	−0.1560	0.251	−0.6211	5.89	.55780
6–10 weeks − 11–16 weeks	−0.2366	0.278	−0.8521	10.74	.41273
Less than 1 week − 11–16 weeks	−0.0172	0.326	−0.0528	14.47	.95858
6–10 weeks − 1–5 weeks	−0.0806	0.200	−0.403	6.51	.69956
Less than 1 week − 1–5 weeks	0.1388	0.262	0.5292	4.65	.62093
Less than 1 week − 6–10 weeks	0.2194	0.288	0.7625	9.18	.46493
Sample size (34 studies; 91 effect sizes)					
More than 150 − 101–150	−0.904	0.428	−2.115	8.75	.0645
1–50 − 101–150	−0.346	0.522	−0.663	11.45	.5205
51–100 − 101–150	−0.662	0.433	−1.528	9.52	.1589934
More than 150 − 1–50	−0.558	0.324	−1.722	9.56	.1171

Comparison A − B	Difference in g	Robust SE	t	df	p
51–100 − 1–50	−0.316	0.331	−0.953	13.49	.3571682
More than 150 − 51–100	−0.242	0.1417	−1.712	7.32	.1288452
Pretest inclusion (34 studies; 91 effect sizes)					
Yes − no	0.352	0.238	1.48	31.7	.1494463
Control group type (34 studies; 91 effect sizes)					
Passive − active	0.0777	0.240	0.324	31.7	.748432

Table D3

AI characteristics

Comparison A − B	Difference in g	Robust SE	t	df	p
Generative AI use (34 studies; 91 effect sizes)					
Yes − no	−0.292	0.223	−1.31	25.6	.2023740
AI application (34 studies; 91 effect sizes)					
ITS/AL − CA	0.321	0.406	0.791	15.01	.441295
LA − CA	−0.511	0.181	−2.825	10.79	.016815
LLM − CA	−0.323	0.242	−1.337	17.86	.197925
LA − ITS/AL	−0.832	0.387	−2.147	11.00	.0549
LLM − ITS/AL	−0.644	0.419	−1.536	14.91	.1455
LA − LLM	−0.188	0.209	−0.896	10.67	.3898
Pedagogical role (34 studies; 91 effect sizes)					
Mixed − feedback/assessment	−0.3962	0.533	−0.744	10.43	.4735
Monitoring/analytics − feedback/assessment	−0.790	0.515	−1.533	6.64	.1715
Scaffolding/facilitation − feedback/assessment	−0.3083	0.532	−0.579	9.88	.575497
Mixed − monitoring/analytics	0.3934	0.1827	2.153	5.51	.0789
Scaffolding/facilitation − monitoring/analytics	0.4812	0.1818	2.647	5.15	.044289
Mixed − scaffolding/facilitation	−0.0879	0.227	−0.387	21.29	.702648

Note. CA = conversational agent; ITS/AL = intelligent tutoring or adaptive learning system; LA = learning analytic system; LLM = LLM-supported system. Mixed denotes the mixed pedagogical-role category.

Table D4

SRL characteristics

Comparison A − B	Difference in g	Robust SE	t	df	p
SRL phase (28 studies; 75 effect sizes)					
Mixed − forethought	0.00654	0.388	0.0168	6.74	.98706
Performance − forethought	0.13723	0.350	0.3924	5.39	.709774
Reflection − forethought	0.29683	0.338	0.8781	6.59	.4108
Mixed − performance	−0.13069	0.273	−0.4785	12.88	.640306
Reflection − performance	0.160	0.296	0.538	6.39	.608519
Mixed − reflection	−0.29029	0.302	−0.9612	8.58	.36277
SRL measure (33 studies; 90 effect sizes)					
Online − offline	−0.230	0.211	−1.09	20.8	.2875361