REVIEW OF CRIME, PEACE AND SOCIETY
ISSN (O): 3079-6105, ISSN (P): 3079-6091
Vol. 3, No. 8, 2026, 375-386

The Impact of Artificial Intelligence on Teaching and Learning Outcomes in Higher Education Institutions: An Integrative Critical Review and Conceptual Framework

Dr. Khadija Awan[1]

Abstract

Artificial intelligence (AI) is rapidly reshaping how higher education institutions (HEIs) design instruction, support learners, and assess achievement. Despite substantial investment and rhetoric, evidence on whether AI improves teaching and learning outcomes remains fragmented across technologies (intelligent tutoring systems, learning analytics, chatbots, and generative AI), disciplines, and outcome constructs. This article presents an integrative critical review of peer-reviewed meta-analyses, systematic reviews, and influential empirical studies, and advances a theory-informed conceptual framework explaining when and why AI affects learning outcomes. Synthesis indicates that (a) intelligent tutoring and adaptive systems yield moderate-to-large positive effects on cognitive achievement relative to conventional instruction, though effects vary with design fidelity, domain, and duration; (b) learning analytics–based early-alert systems improve retention and course performance mainly when coupled with human intervention; (c) generative AI shows promising but heterogeneous effects, with strong dependence on pedagogical scaffolding and risk of cognitive offloading when used as an answer engine; and (d) faculty-facing AI can reduce routine workload and enhance feedback timeliness but raises concerns regarding academic integrity, algorithmic bias, privacy, and professional agency. The proposed Pedagogically-Mediated AI Impact (PMAI) framework positions pedagogy, learner characteristics, institutional capacity, and governance as mediating and moderating conditions between AI affordances and outcomes. Eight testable propositions and a research agenda are offered. The article concludes that AI is neither a panacea nor a threat in itself; its educational value is contingent on intentional pedagogical integration, ethical governance, and sustained faculty development.

Article History
Received 22 September 2026
Accepted 14 October 2026
Published 20 October 2026
Corresponding Author*
OPEN ACCESS
Keywords
Artificial intelligence in education; higher education; learning outcomes; intelligent Tutoring systems; Learning analytics; generative AI; academic Integrity; pedagogy

1. Introduction

Higher education is undergoing a technological transition comparable in significance to the introduction of the learning management system, yet moving considerably faster. Artificial intelligence, broadly defined as computational systems that perform tasks associated with human cognition such as perceiving, reasoning, predicting, and generating language, has moved from specialised research laboratories into everyday academic practice (Holmes et al., 2019; Zawacki-Richter et al., 2019). Students now routinely interact with adaptive platforms, automated feedback tools, and large language model (LLM) chatbots, while institutions deploy predictive analytics to identify students at risk of attrition (Viberg et al., 2018; Kasneci et al., 2023).

The promise of AI in education is longstanding. Bloom's (1984) "2 sigma problem" demonstrated that one-to-one tutoring could raise average achievement by two standard deviations over conventional classroom teaching, and subsequent research on intelligent tutoring systems (ITS) has sought to approximate this effect at scale (VanLehn, 2011). More

[1], V. Principal, Allied Schools, Sapphire Campus, Zafar ul Haq road, Rawalpindi. Email: Khadija.awan24@gmail.com

recently, the public release of generative AI tools has dramatically lowered the barrier to adoption and intensified debate about the nature of learning, authorship, and assessment in universities (Dwivedi et al., 2023; Cotton et al., 2024).

Despite this momentum, three problems limit evidence-based decision-making. First, the empirical literature is dispersed across technologies and outcome measures, making generalisation difficult. Bond et al. (2024), in a meta systematic review of AI in higher education, observed limited ethical engagement, weak theoretical grounding, and insufficient involvement of educators in much published work. Second, a large share of studies privileges short-term perceptions and technology acceptance over robust measures of learning, such as transfer, retention, and higher-order thinking. Third, critical scholarship warns that AI adoption may be driven by commercial and managerial logics rather than pedagogical need (Selwyn, 2019; Williamson & Eynon, 2020).

This article addresses these gaps by synthesising existing high-quality evidence and proposing an integrative framework. It is guided by the following research questions:

- **RQ1.** What is the evidence regarding the effect of major categories of AI applications on student learning outcomes in higher education?
- **RQ2.** How is AI changing teaching practice, assessment, and the role of the instructor?
- **RQ3.** Which pedagogical, learner-related, and institutional conditions moderate the relationship between AI use and learning outcomes?
- **RQ4.** What ethical and governance challenges must be addressed to ensure that AI improves rather than undermines educational quality?

The contribution is threefold: a cross-technology synthesis of outcome evidence, a theoretically grounded framework (PMAI) that explains heterogeneity in results, and a set of propositions to direct future empirical work.

2. Theoretical Background

2.1 Defining AI and learning outcomes in higher education

AI in education (AIEd) encompasses student-facing systems (adaptive tutors, conversational agents, writing assistants), instructor-facing systems (automated grading, content generation, dashboards), and institution-facing systems (admission, retention prediction, resource allocation) (Zawacki-Richter et al., 2019; Chen et al., 2020). *Learning outcomes* are conceptualised here along four dimensions: (a) cognitive achievement (knowledge, grades, test performance); (b) higher-order skills (critical thinking, problem solving, creativity); (c) affective and motivational outcomes (engagement, self-efficacy, satisfaction); and (d) progression outcomes (persistence, retention, completion).

2.2 Learning theories relevant to AI-mediated instruction

Three theoretical lenses are particularly useful. **Cognitive load theory** posits that learning is constrained by limited working memory and that instruction should manage intrinsic, extraneous, and germane load (Sweller, 1988). Adaptive systems can reduce extraneous load and match task difficulty to learner expertise, whereas poorly designed generative tools may reduce the effortful processing necessary for durable learning. **Sociocultural theory**

emphasises learning as mediated by tools and more capable others within the zone of proximal development (Vygotsky, 1978). AI systems can function as scaffolding agents, yet their effectiveness depends on whether support is progressively faded. **Self-regulated learning (SRL)** theory highlights goal setting, monitoring, and strategic adaptation (Zimmerman, 2002); AI-generated analytics and feedback can strengthen metacognitive monitoring, but over-reliance may displace learners' own regulation.

At the teacher level, the technological pedagogical content knowledge (TPACK) framework argues that technology integration is effective only when instructors combine technological, pedagogical, and content expertise (Mishra & Koehler, 2006). Adoption behaviour is further explained by the technology acceptance model and the unified theory of acceptance and use of technology, which identify perceived usefulness, ease of use, social influence, and facilitating conditions as drivers of use (Davis, 1989; Venkatesh et al., 2003). Together, these frameworks suggest that outcomes depend not on technology alone but on the interplay of learner cognition, instructional design, and institutional context.

3. Method

3.1 Review design

This study adopts an integrative critical review design, appropriate when the aim is to synthesise heterogeneous evidence and generate theory rather than to pool effect sizes (Tranfield et al., 2003). The approach prioritises second-order evidence (meta-analyses and systematic reviews) supplemented by landmark primary studies, which allows conclusions to rest on cumulative rather than isolated findings. Reporting follows the transparency principles of the PRISMA 2020 statement where applicable (Page et al., 2021).

3.2 Eligibility and analytic strategy

Sources were considered eligible if they (a) were peer-reviewed or published by recognised scholarly or policy bodies; (b) examined AI-enabled tools in post-secondary contexts, or generalisable findings from adjacent contexts; and (c) reported or discussed learning, teaching, or institutional outcomes. Evidence was organised thematically using an iterative coding logic consistent with thematic analysis (Braun & Clarke, 2006), resulting in four technology clusters (intelligent tutoring and adaptive systems, learning analytics and prediction, conversational and generative AI, and automated assessment and feedback) and four cross-cutting themes (pedagogical integration, equity and bias, academic integrity, and governance).

3.3 Note on evidentiary limits

The pace of generative AI development means that primary evidence is accumulating faster than it can be consolidated. Findings concerning LLM-based tools are therefore treated as provisional. Authors adapting this manuscript for submission should re-run a documented database search (e.g., Web of Science, Scopus, ERIC) covering the most recent period, report the full search string, screening counts, and quality appraisal, and update the synthesis accordingly.

4. Synthesis of Evidence

4.1 Intelligent tutoring systems and adaptive learning

The most mature evidence base concerns ITS. VanLehn (2011) reviewed comparisons of human tutoring, ITS, and other tutoring and concluded that step-based ITS were nearly as effective as human tutors, with effect sizes substantially above those of no tutoring. Meta-analyses focusing on college students corroborate this pattern: Steenbergen-Hu and Cooper (2014) reported a moderate positive effect of ITS on college students' academic learning, and Ma et al. (2014) found that ITS outperformed large-group human instruction, individualised computer-based instruction, and textbook or workbook use, with a mean effect size around moderate magnitude. Kulik and Fletcher (2016) reviewed controlled evaluations and reported a median effect size in the moderate-to-large range, noting that effects were larger for locally developed systems and evaluations using local rather than standardised tests, a caution regarding measurement alignment.

More recent meta-analyses on broader AI applications confirm positive but variable effects. Zheng et al. (2023) synthesised studies on AI and learning achievement and perception, finding overall positive effects moderated by factors such as intervention duration, learning environment, and the role AI played in instruction. Collectively, this literature suggests that the *mechanism* of impact is real-time diagnostic feedback, mastery-based sequencing, and reduced waiting time for help, all consistent with established principles of effective instruction (Hattie, 2009).

Importantly, ITS benefits are not uniform. They tend to be strongest in well-structured domains (mathematics, programming, physics, language grammar) where knowledge components can be modelled explicitly, and weaker in ill-structured domains requiring interpretation, argumentation, or creativity. Furthermore, a recurrent critique is that laboratory or short-term quasi-experiments overstate effects achievable under routine institutional conditions.

4.2 Learning analytics and predictive systems

Learning analytics (LA) uses data from learner interactions to understand and optimise learning and its environments (Siemens & Long, 2011). Applications range from descriptive dashboards to predictive models of attrition and prescriptive recommendations. Reviews of LA in higher education show growth in implementation but caution that many initiatives remain at pilot stage, with limited evidence of impact on learning itself (Viberg et al., 2018). Gašević et al. (2015) argued that analytics must be grounded in learning science, lest institutions optimise what is easily measured rather than what matters.

A frequently cited institutional example is the Course Signals project at Purdue University, which used predictive modelling to provide students with traffic-light feedback and reported higher retention among participating cohorts (Arnold & Pistilli, 2012); however, subsequent methodological critiques questioned causal claims because of selection effects, illustrating the importance of rigorous quasi-experimental or randomised designs. The more general lesson is that the predictive accuracy of a model is insufficient: outcomes improve when predictions

trigger timely, personalised human outreach such as advising, tutoring, or instructor contact. Analytics without intervention yields information, not improvement.

4.3 Chatbots and generative AI

Conversational agents and LLM-based tools have rapidly diffused through universities. Wu and Yu (2024) conducted a meta-analysis of AI chatbot interventions and found a positive overall effect on learning outcomes, with the magnitude varying by intervention length, educational level, and the pedagogical role of the chatbot. Emerging syntheses on ChatGPT specifically report positive average effects on achievement, learning perception, and some higher-order thinking indicators, though the primary studies are mostly short-term, heterogeneous in design, and at risk of novelty effects (Deng et al., 2025). Kasneci et al. (2023) outline potential benefits, including personalised explanations, support for writing and language learning, and automated generation of practice items, alongside risks, including inaccuracy, bias, and over-reliance.

A crucial distinction emerges between **AI as tutor** and **AI as answer engine**. When systems are configured to prompt reasoning, offer hints rather than solutions, and require learner production, they align with constructivist and SRL principles and plausibly promote durable learning. When students outsource the cognitively demanding portions of tasks, short-term performance may rise while underlying skill acquisition stagnates, a phenomenon often termed *cognitive offloading* or the "illusion of competence." This interpretation is consistent with cognitive load theory's emphasis on germane processing (Sweller, 1988) and underlines why unrestricted access does not necessarily translate into learning gains.

4.4 Automated assessment and feedback

Timely, specific feedback is among the most powerful influences on achievement (Hattie, 2009). AI-enabled automated essay scoring, code evaluation, and feedback generation can expand the volume and speed of feedback, particularly in large-enrolment courses where instructor feedback is scarce. Evidence indicates acceptable agreement between automated and human scoring for structured tasks, but weaker validity for creativity, argument quality, and discipline-specific nuance (Zawacki-Richter et al., 2019). Feedback effectiveness also depends on whether students understand and act on it; thus, AI-generated feedback should be embedded in cycles of revision and reflection rather than delivered as a terminal evaluation.

4.5 Impact on teaching practice and the instructor's role

AI is reshaping instructor work in at least three ways. First, it automates routine tasks (drafting materials, generating quizzes, grading objective items, answering frequently asked questions), potentially freeing time for mentoring and higher-value interaction (Holmes et al., 2019). Second, it provides instructors with data to adjust teaching, for example by identifying common misconceptions. Third, it shifts expectations about professional competence, requiring "AI literacy" alongside disciplinary and pedagogical expertise (Mishra & Koehler, 2006; Luckin et al., 2016).

However, workload relief is not guaranteed. Verification of AI outputs, redesign of assessments, and management of integrity concerns may introduce new labour. Critical

scholars also warn of deskilling and surveillance risks if analytics are used to monitor instructors, and of the displacement of professional judgment by opaque systems (Selwyn, 2019; Williamson & Eynon, 2020). Luckin et al. (2016) advocate a complementary model in which AI handles what machines do well, while educators concentrate on empathy, ethical reasoning, and complex facilitation.

5. Moderators and Mediators of AI Impact

The heterogeneity in results across studies is not random; it reflects systematic conditions.

Pedagogical design. Studies repeatedly show that AI effects are larger when integrated into well-structured instructional designs, such as active learning, formative assessment, and mastery learning, than when added as a standalone resource (Freeman et al., 2014; Hattie, 2009). AI amplifies the pedagogy within which it is embedded.

Learner characteristics. Prior knowledge, digital literacy, motivation, and self-regulation influence benefits. Learners with strong SRL skills tend to use AI strategically, whereas novices may accept outputs uncritically. This raises an equity concern: AI could widen gaps if advantaged students leverage it more effectively (Zimmerman, 2002).

Task and domain. Well-defined tasks benefit more readily from adaptive modelling; open-ended tasks require careful design to prevent substitution of student thinking.

Duration and novelty. Short interventions often show larger effects than longer ones, partly because of novelty and partly because of tight alignment with immediate assessments (Kulik & Fletcher, 2016; Wu & Yu, 2024). Longitudinal evidence is scarce.

Institutional capacity. Infrastructure, faculty training, technical support, and leadership commitment shape implementation quality. Adoption theory predicts that facilitating conditions and perceived usefulness drive sustained use (Venkatesh et al., 2003).

Faculty readiness. TPACK suggests that instructors who can align AI affordances with disciplinary pedagogy produce stronger outcomes than those who simply permit or prohibit tools (Mishra & Koehler, 2006).

6. Risks, Ethics, and Governance

6.1 Algorithmic bias and fairness

Machine-learning models trained on historical data can reproduce or amplify inequities related to gender, ethnicity, socioeconomic status, or language background (Mehrabi et al., 2021). In higher education, predictive models that label certain student groups as high-risk may lead to lower expectations or stigmatisation. Fairness audits, transparent model documentation, and inclusion of diverse stakeholders in design are therefore essential.

6.2 Privacy and data ethics

AI-based systems depend on extensive collection of behavioural and performance data. Slade and Prinsloo (2013) identified core ethical tensions in learning analytics: informed consent, student agency, the right to opt out, data ownership, and the duty to act on knowledge gained.

These considerations are heightened for commercial generative AI platforms, in which prompts and uploaded materials may be processed by external providers. Institutions require clear data-governance policies, vendor due diligence, and transparent communication with students.

6.3 Academic integrity and authentic assessment

Generative AI has intensified concerns about plagiarism and contract-cheating-like practices (Cotton et al., 2024). Detection-based approaches are unreliable and can produce false accusations, particularly against non-native English writers. A more sustainable response is assessment redesign: process-oriented tasks, oral defences, in-class components, reflective commentary, and explicit guidance on legitimate AI use. Integrity policy should shift from prohibition alone to the cultivation of ethical, transparent AI practice.

6.4 Accuracy, transparency, and epistemic risk

LLMs may generate plausible but incorrect statements and fabricated citations, creating epistemic hazards when students lack the expertise to detect errors (Kasneci et al., 2023; Dwivedi et al., 2023). Developing critical evaluation skills is thus a core learning outcome in its own right.

6.5 Commercialisation and the digital divide

Critical perspectives argue that AIEd is shaped by market incentives and may deepen dependence on private vendors (Williamson & Eynon, 2020). Access disparities, including connectivity, devices, and subscription costs, can reproduce inequalities, particularly in resource-constrained institutions and developing regions.

7. The Pedagogically-Mediated AI Impact (PMAI) Framework

Synthesising the evidence, this article proposes the PMAI framework. Its central claim is that AI affordances (personalisation, immediacy, scalability, generativity, and prediction) do not directly cause improved outcomes; rather, their effects are *mediated* by the quality of pedagogical integration and the nature of student cognitive engagement, and *moderated* by learner, task, and institutional conditions, all within an ethical-governance envelope.

Components of the framework:

1. **AI affordances** (input): personalised pathways, real-time feedback, predictive insight, content generation, conversational support.
2. **Pedagogical mediation** (mediator 1): alignment with learning objectives, scaffolding and fading, formative assessment cycles, and promotion of active rather than passive use.
3. **Cognitive and behavioural engagement** (mediator 2): germane cognitive load, SRL strategies, time on task, and critical evaluation of AI output.
4. **Moderators**: learner prior knowledge and AI literacy; task type; intervention duration; faculty TPACK; institutional infrastructure and culture.
5. **Governance envelope**: fairness, privacy, transparency, integrity policy, and human oversight.

6. **Outcomes**: cognitive achievement, higher-order skills, motivation and self-efficacy, retention, and equity of outcomes.

7.1 Propositions

- **P1.** AI-enabled adaptive instruction will improve cognitive achievement relative to non-adaptive instruction, with stronger effects in well-structured domains.
- **P2.** The effect of AI on achievement is mediated by the degree of pedagogical alignment; AI added without redesign will yield negligible or inconsistent gains.
- **P3.** AI configured to elicit learner reasoning (hints, questioning) will produce greater delayed-retention and transfer outcomes than AI configured to provide complete solutions.
- **P4.** Predictive analytics will improve retention only when coupled with timely human intervention.
- **P5.** Students' prior knowledge and self-regulation positively moderate the benefits of generative AI, potentially widening achievement gaps absent targeted support.
- **P6.** Faculty AI-pedagogical competence positively moderates the relationship between AI adoption and learning outcomes.
- **P7.** Perceived fairness and transparency of AI systems positively predict student trust, engagement, and willingness to use AI appropriately.
- **P8.** Assessment redesign emphasising process and authentic performance reduces integrity violations more effectively than detection-based policies.

8. Implications

8.1 Implications for theory

The PMAI framework contributes by shifting analytic attention from "does AI work?" to "under what conditions, for whom, and through which mechanisms does AI work?" This reframing integrates cognitive load theory, SRL, TPACK, and technology adoption models into a single explanatory structure and offers a basis for moderated-mediation modelling in future studies.

8.2 Implications for institutional leaders and policymakers

Institutions should adopt AI strategically rather than reactively. Recommended actions include: (a) developing an institution-wide AI strategy aligned with teaching and learning priorities rather than efficiency alone; (b) investing in sustained, discipline-specific faculty development; (c) establishing governance bodies including students and educators to review tools for bias, privacy, and pedagogical value; (d) ensuring equitable access to approved tools; and (e) building evaluation mechanisms to monitor impact on learning rather than usage statistics.

8.3 Implications for instructors

Instructors should define explicit, course-level AI-use policies; design tasks that require reasoning, justification, and reflection; teach students to verify and critique AI outputs; and use analytics to inform, not replace, professional judgment. Modelling transparent AI use can further promote academic integrity and digital literacy.

8.4 Implications for students

Students benefit from developing AI literacy, including understanding system limitations, prompting for explanation rather than answers, and documenting AI assistance. Metacognitive habits, such as attempting problems before consulting AI and testing understanding without assistance, can protect against offloading.

9. Limitations and Future Research

Several limitations qualify this review. First, as an integrative review emphasising second-order evidence, it inherits the quality limits of included syntheses, including publication bias and heterogeneity in outcome measures. Second, the rapid evolution of generative AI means that conclusions about specific tools may become outdated quickly. Third, the literature is geographically skewed toward a limited number of countries, which constrains generalisability to diverse institutional and cultural contexts, including those in South Asia and other developing regions. Fourth, the PMAI framework is conceptual and requires empirical validation.

Future research should: (1) conduct pre-registered randomised or rigorous quasi-experimental studies with delayed post-tests to assess retention and transfer; (2) examine heterogeneity of effects across demographic and disciplinary groups using fairness-aware designs; (3) investigate long-term effects on critical thinking, creativity, and professional competence; (4) test moderated-mediation models derived from PMAI using structural equation modelling or multilevel designs; (5) explore faculty experiences, workload, and professional identity through mixed-methods and longitudinal designs; (6) evaluate the cost-effectiveness of AI initiatives relative to alternative investments; and (7) develop and validate instruments for AI literacy and AI-supported learning behaviours. Greater theoretical grounding, educator involvement, and ethical reflection, as called for by Bond et al. (2024), should characterise this work.

10. Conclusion

The evidence reviewed indicates that AI can improve teaching and learning outcomes in higher education, but its effects are conditional rather than automatic. Intelligent tutoring and adaptive systems provide the strongest evidence for gains in cognitive achievement; learning analytics improve progression when connected to human action; and generative AI offers considerable potential if designed and used to stimulate rather than replace student thinking. At the same time, risks associated with bias, privacy, integrity, epistemic error, and commercial dependence demand robust governance. The PMAI framework advanced here emphasises that pedagogy, learners, faculty, and institutional context determine whether AI becomes a catalyst for deeper learning or a shortcut that erodes it. For higher education institutions, the central task is therefore not simply to adopt AI, but to integrate it deliberately, evaluate it rigorously, and govern it ethically, keeping human educators and learning itself at the centre.

Declarations

Funding: [To be completed by authors.] **Conflict of interest:** [To be completed by authors.] **Ethics statement:** This study is a review of published literature and involved no human participants. **Data availability:** No new datasets were generated.

References

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

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16. https://doi.org/10.3102/0013189X013006004

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, Article 4. https://doi.org/10.1186/s41239-023-00436-z

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

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*, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918. https://doi.org/10.1037/a0037123

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), Article 115. https://doi.org/10.1145/3457607

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 30–40.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology, 106*(2), 331–347. https://doi.org/10.1037/a0034752

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management, 14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology, 55*(1), 10–33. https://doi.org/10.1111/bjet.13334

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*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

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

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2