OPEN ACCESS

Saudi Journal of Engineering and Technology
Abbreviated Key Title: Saudi J Eng Technol
ISSN 2415-6272 (Print) | ISSN 2415-6264 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Original Research Article

AI for Everything: Opportunities, Challenges, and Responsible Use of Artificial Intelligence in School Education

Layla Sultan[1]*, Abdulla Hamad Alali[2], Saif Hamad Alali[3], Zayed Humaid Yousuf[4]

[1]Senior Manager Equipment Commissioning, DEWA, UAE
[2]High School Diploma Student at International Schooling, Dubai, UAE
[3]Student in GEMS School of Research and Innovation, Dubai, UAE
[4]Student GEMS International School – Dubai Hills (Al Khail), Dubai, UAE

DOI: https://doi.org/10.36348/sjet.2026.v11i11.003 | **Received:** 05.09.2026 | **Accepted:** 29.10.2026 | **Published:** 02.11.2026

***Corresponding author:** Layla Sultan
Senior Manager Equipment Commissioning, DEWA, UAE

Abstract

AI assistance can improve a pupil's immediate output without improving independent learning. This targeted, structured review examines school education at approximately ages 6–18, synthesizing 25 cited sources and extracting comparable design and outcome information from eight studies or reviews. Logged revision searches, DOI metadata, source decisions, and extraction records support auditability without claiming a comprehensive systematic search. Domain-specific intelligent tutoring systems and general-purpose generative AI have different evidence bases. Effects depend on comparators, outcome measures, implementation, and assessment conditions; confidence intervals and reported uncertainty are retained rather than combining incompatible results. A school evidence table distinguishes pupils from repeated observations and immediate assisted performance from verified unaided outcomes. Self-regulated learning, cognitive offloading, and scaffolding inform a proposed attempt–consult–verify–explain–fade–transfer routine. An unvalidated decision aid links developmental context, tool eligibility, supervision, and local unaided-learning evidence, complementing existing competency, child-rights, and assessment-regime frameworks. Schools should prioritize learning, inclusion, safety, and teacher responsibility, and revise or stop uses that do not meet those goals.

Keywords: Artificial intelligence; generative AI; school education; AI literacy; personalized learning; responsible AI.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

1. INTRODUCTION

A student who receives an instant explanation may feel less stuck. A teacher who can adapt a reading passage quickly may have more time to discuss it. These possibilities help explain schools' interest in artificial intelligence, but neither convenience nor fluency proves educational value. The central question is whether an application helps students understand, practise, and eventually act without it. UNESCO's human-centred guidance places educational purposes and human agency ahead of technological expansion (Miao & Holmes, 2023). Figure 1 illustrates this priority: technology accompanies conversation, books, and student work rather than replacing them.

AI refers here to computer systems that perform tasks such as recognizing patterns, making predictions, or producing recommendations that ordinarily require aspects of human intelligence. In education, an intelligent tutoring system typically adapts tasks and feedback within a defined subject. Generative AI is different: it creates content, including text, images, or code. A large language model generates text from patterns learned during training; an apparently persuasive answer is not a guarantee of accuracy (Kasneci *et al.,*2023).

AI literacy involves understanding, using, and critically evaluating AI while recognizing its ethical and social implications (Long & Magerko, 2020; Ng *et al.,*2021). It is broader than writing effective prompts. A literate learner can ask where an answer came from, identify an unsupported claim, protect personal information, and decide when a task should remain unaided. For children, these judgments develop alongside subject knowledge, self-regulation, and

relationships with adults. Consequently, access should not be mistaken for readiness. Effective school policy must accommodate these differences rather than applying the same expectations to all learners.



Figure 1: Teacher-guided AI exploration. Original AI-generated conceptual illustration, not participant data

The phrase 'AI for Everything' invites examination across school education, not use in every activity. Three questions guide the review: which benefits are demonstrated, which risks require attention, and how should schools decide when assistance is appropriate? The contribution is a proposed decision aid linking age and eligibility, bounded ITS versus general GenAI, supervision, and demonstrated unaided learning. It does not claim priority for separating assisted performance from learning (Kovari, 2026). Figure 2 adds an operational school-decision layer to that distinction: teachers retain responsibility, students verify and explain, and failed learning or safety checks trigger revision or stopping.

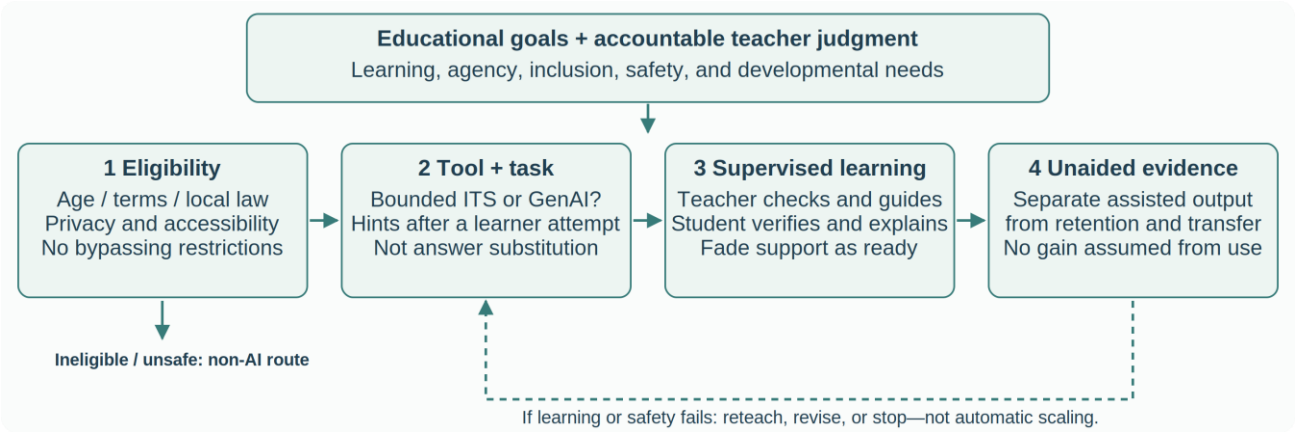


Figure 2: Proposed decision aid linking goals, eligibility, tool design, teacher supervision, and unaided learning. Conceptual synthesis, not a validated age rule

2. METHODOLOGY

This is a targeted structured review and conceptual synthesis, not a systematic or comprehensive review. Citation-led retrieval began on 18 October 2026; revision searches and verification occurred on 19 October 2026 (UTC+4; 18 October UTC). Manus Search, in scholar and web modes, located publications and official guidance; public-page extraction retrieved publisher, repository, and organizational text. Crossref's REST works endpoint checked 20 DOI metadata records, with responses archived. No direct Scopus, Web of Science, ERIC, or Google Scholar database search is claimed.

Three logged discovery queries returned 46 result records, representing 45 distinct URLs. Supplement S2 preserves their exact wording, service, date, returned records, and decisions; item-level author/DOI queries are also recorded. Eligibility favoured peer-reviewed comparative school learning evidence, relevant tutoring reviews, foundational learning theory, and authoritative child/education guidance. Cross-sector findings are explicitly qualified; guidance is normative, not efficacy data. The predefined extraction set comprises three comparator-sensitive historical meta-analyses, four school empirical studies, and one recent K–12 review. Twenty documented sources from the previous version were retained; five

© 2026 | Published by Scholars Middle East Publishers, Dubai, United Arab Emirates

theory/review/framework sources were added, yielding 25 citations.

Duplicate handling used exact URLs, verified DOIs, and known publication/report locators, preserving alternative versions and prioritizing current published results. Social/promotional or secondary commentary was not advanced as primary evidence. Other discoveries outside the purposive corpus did not receive exhaustive full-text eligibility assessment; their possible relevance is not denied. Historical complete search-result sets and screening counts were unavailable and were not retrospectively invented. The archived snapshot permits audit, although dynamic ranking and unspecified underlying index coverage limit exact future retrieval reproduction.

AI-assisted extraction recorded population, unique versus repeated-observation counts, tool, design, comparator, exposure, outcome, uncertainty, location, and limitations. A descriptive five-domain appraisal covered selection/allocation, measurement alignment, missing data, implementation confounding, and generalizability; it is not a validated risk-of-bias score. Coding distinguished self-regulation, offloading, scaffolding, assisted versus unaided outcomes, and governance. No independent dual-human screening, reviewer-agreement statistic, original participant study, or new effect pooling was performed. Selective English-language retrieval limits coverage; the proposed diagrams and routine remain unvalidated.

3. ANALYSIS AND DISCUSSION

3.1 Comparative evidence and quantitative interpretation

Historical ITS effects are comparator-sensitive. Ma et al. (2014) reported random-effects g=.41, 95% CI [.34, .48], across 107 effect-size units and 14,321 participants, not a school-only pool. Kulik and Fletcher (2016) reported weighted g=.50 [.40, .59] versus conventional instruction; this differs from their unweighted median Glass's effect size of .66. In K–12 mathematics, Steenbergen-Hu and Cooper (2013) found random-effects g=.09 [−.03, .20] unadjusted and .01 [−.10, .12] adjusted. Overlapping evidence, test alignment, and design differences preclude averaging these reviews. Step-based tutoring also differs from answer marking (VanLehn, 2011).

Table 1 compares school evidence. Bastani's Base exam coefficient −.054 (SE .022) is a 17% relative reduction against the .321 control mean, not 17 percentage points; the Tutor estimate −.004 (SE .013) is not evidence of equivalence. Fütterer's active-AI comparator cannot estimate AI versus no-AI teaching, while substantial attrition weakens a null inference. Yang's quasi-experimental reporting inconsistencies warrant caution. Except where explicitly established, posttests are not relabelled unaided. Standardized effects, normalized coefficients, test statistics, and review direction counts are incompatible metrics; none establishes durable benefit across ages 6–18.

Table 1: Comparative school AI evidence: studies, outcomes, and limits

Study / tool	Population / sample	Design / comparator	Exposure	Outcome / estimate	Limits
Roschelle *et al.,* (2016) ASSISTments + training	2,850 pupils; 43 Maine Grade-7 schools	Cluster RCT; full package versus usual homework	School year; 14 h mean use	TerraNova: adjusted +8.835 points; g=.18, p=.007	10.1% attrition; package effects; CI/SE and test AI access NR
Bastani *et al.,* (2025) GPT-4 Base / Tutor	Turkish Grades 9–11; nearly 1,000 pupils; 2,848 student-session observations	Three-arm classroom RCT; Base or Tutor versus books/notes	4 × 90 min; next exam	0–1 score: Base practice +.137 (SE .031), unaided exam −.054 (SE .022); Tutor +.361 (SE .032), −.004 (SE .013)	Exact unique N/CI/attrition unverified; one school; immediate near-task outcome
Fütterer *et al.,* (2026) Custom GPT-4o	371 pupils; German Grades 7–9 physics/English	Individual RCT; pedagogical prompts versus active AI	4 × 45 min lessons; next test	Knowledge: $\chi^2(1)=1.28$, p=.257; numeric CI / standardized effect NR	35% dropout; active comparator; test AI access NR
Yang *et al.,* (2025) ChatGPT for C++	153 pupils; 6 classes; Taiwan girls' senior high school	Quasi-experiment; add-on AI versus conventional teaching; pretests controlled	6 weeks; AI last 3; next test	Achievement (Table 3): F=3.96, partial η^2=.03, p=.048; AI group reported lower	Allocation/clustering unclear; group-N and F reporting conflicts; AI-free test NR
Létourneau *et al.,* (2025) Diverse ITS	28 studies / 26 reports; aggregate N=4,597, not pooled-effect N	Systematic review; varied comparators; no meta-analysis	One class to 30 weeks; half <1 week	7/8 versus usual teaching reported significant benefit; 1/4 versus non-intelligent systems did	Heterogeneous; direction counts are not pooled efficacy; common unaided/delayed status unknown

g = Hedges' g; SE = standard error; CI = confidence interval; NR = not reported in verified text; RCT = randomized trial; ITS = intelligent tutoring system. Bastani practice p<.01; Base exam p<.05; Tutor exam not significant. Metrics are not pooled. Details/corrections: Supplement S2 and workbook.

3.2 Learning mechanisms: a conditional explanation

Self-regulated learning organizes planning, performance monitoring, and reflection (Panadero, 2017). The proposed routine begins with an independent attempt, then consultation, verification against subject evidence, and an explanation in the learner's own words. These steps expose goals, uncertainty, and strategy choice rather than rewarding completion alone. Assistance is then reduced, and a related task is attempted without AI; delayed transfer should be assessed separately.

Cognitive offloading changes a task's information-processing demands through an external aid (Risko & Gilbert, 2016). A response may free capacity for comparing explanations, or replace the calculation, retrieval, or reasoning being learned. Scaffolding offers a conditional design principle: assistance should respond to the learner's difficulty rather than supply every answer (Wood *et al.,*1976). Bounded ITS can constrain help; general GenAI more readily supplies a finished artefact. These are inferred mechanisms, not demonstrated AI-mediated causal pathways. Researchers should measure attempt quality, consultation timing, verification, explanation, and fading alongside independent outcomes, rather than infer mechanisms from engagement.

3.3 Development, subject applications, and AI literacy

Age is a decision context, not a proxy for competence. Table 2 offers progressively greater responsibility while retaining adult support. Primary pupils can explore patterns and model errors without personal accounts; adolescents can compare sources and defend revisions. UNESCO's recommendation of 13 for classroom AI use is guidance, not a universal legal rule or a demonstrated readiness threshold (UNESCO, 2023). Current product terms, local law, guardian requirements, and accessibility needs must be checked; adult mediation must not bypass restrictions or upload children's data. UNICEF (2025) grounds these choices in safety, agency, development, and inclusion.

Table 2: Age-appropriate AI use

School stage	Suitable activities	Supervision	Safeguards
Primary, approximately 6–11	Paper-based AI literacy; teacher demonstrations	Adult operates approved tools	No personal accounts or identifiable data; retain play and peer interaction
Lower secondary, approximately 12–14	Compare explanations; evaluate model errors	Guided, bounded tasks; verify eligibility	Check age terms; teach privacy and source checking
Upper secondary, approximately 15–18	Critique drafts; test code; document assistance	Teacher sets permitted uses and reviews reasoning	Protect data; disclose assistance; assess unaided work

Unvalidated stage guidance; boundaries and eligibility vary. UNESCO (2023), Miao et al. (2024), and UNICEF (2025).

Figure 3 and Supplement S1 illustrate, rather than evaluate, subject applications. Mathematics hints follow an attempt; science claims require evidence and safety checks; languages retain conversation and independent writing. Humanities tasks verify sources, arts discuss attribution, and computing requires explanation and testing of suggested code. AI literacy means understanding and evaluating fallible outputs, including knowing when not to use them (Long & Magerko, 2020; Miao *et al.,*2024).



Figure 3: Supervised subject activities. Original AI-generated conceptual illustration, not research findings

3.4 Teacher judgment and inclusion

Teachers may draft lessons or feedback with AI, but must check accuracy, curriculum, culture, and accessibility. Consequential grading and placement remain accountable human decisions. UNESCO's teacher competencies address pedagogy and ethics, not

delegated responsibility (Miao & Cukurova, 2024). Teacher-authored GPT Tutor safeguards required substantial work (Bastani *et al.,*2025); workload evaluation must include checking, corrections, preparation, and incidents rather than assume time savings.

Assistive speech, translation, or adaptive presentation is not necessarily GenAI. Disability-intervention evidence is promising but methodologically uncertain; none of the studies in Paglialunga and Melogno's (2025) review was rated low risk of bias, and its population extended beyond school. Trial support with intended users and specialists, retain accessible alternatives, and examine accent, language, device, and cost barriers. Inclusion requires usable access and a voice in decisions, not merely log-in permission.

3.5 Risks, integrity, and well-being

Fluent outputs can contain errors, fabricated sources, or stereotypes. Upstream approval and data minimization matter more than warnings alone (Miao & Holmes, 2023). Table 3 and Figure 4 distinguish potential benefits from safeguards. A polished submission can hide missing practice, so permitted assistance, drafts, explanation, and appropriately designed no-AI work should be specified before assessment.

Automated suspicion is not proof of misconduct. Students need a safe reporting route and access to trusted adults; a chatbot should not replace counselling or relationships. UNICEF's (2025) updated child-rights guidance explicitly addresses generative AI and companions. Safety, privacy, and fairness checks are necessary even when a learning outcome improves.

Table 3: Benefits, risks, and safeguards

Potential benefit	Associated risk	Evidence or limitation	Practical response
Timely hints	Answer copying	Bastani *et al.,* (2025): practice and unaided outcomes diverged	Ask for attempts; give hints; retest
Adaptive practice	Uneven benefits	K–12 effects are context-dependent	Monitor progress by learner group
Accessible materials	New access barriers	Disability evidence has bias concerns	Test with users; retain alternatives
Faster preparation	Incorrect materials	Commentary is not workload evidence	Review accuracy; log checking time
Dialogue support	Privacy or relational harm	Child-rights guidance, not efficacy data	Minimize data; retain adult support
Draft feedback	Hidden outsourcing	Assisted output does not prove skill	Disclose use; inspect drafts and reasoning

Conceptual safeguards, not tested interventions. Sources: Sections 3.4–3.5.

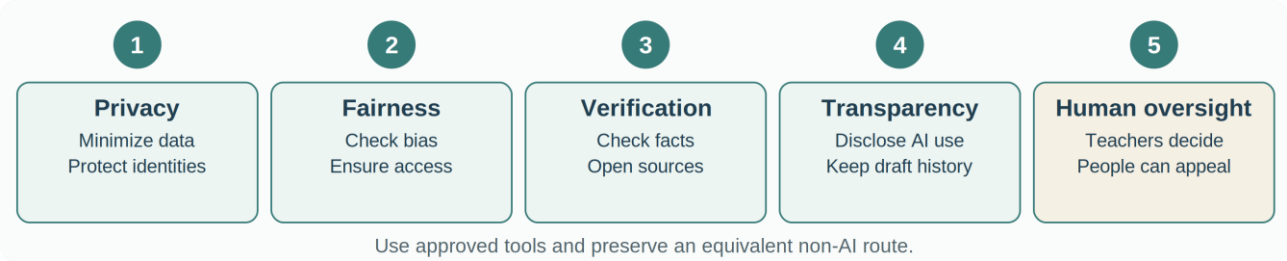


Figure 4: Privacy, fairness, verification, transparency, and human oversight. Conceptual UNESCO/UNICEF synthesis

3.6 What the synthesis adds: decision rules

The decision aid adds a school-operational layer, not a claim that planning and piloting are new. UNESCO's student framework specifies 12 competencies across four dimensions and its teacher framework 15 across five; UNICEF supplies ten child-rights requirements (Miao *et al.,*2024; Miao & Cukurova, 2024; UNICEF, 2025). Kovari (2026) already distinguishes assisted performance, independent performance, retention, and transfer in assessment-regime reporting. Table 4 complements these references by linking developmental context and access eligibility to tool design, required supervision, and an unaided-learning decision. Competencies inform preparation, rights constrain eligibility, and outcome regimes guide evaluation. Neither the matrix nor its stop/scale criteria are validated or universal.

Use the routine only for an approved task: attempt first; consult for limited help; verify; explain; fade assistance; demonstrate unaided transfer. Source checks should suit the discipline, and accommodations should preserve access without outsourcing the skill being assessed. A failure to explain or transfer triggers

© 2026 | Published by Scholars Middle East Publishers, Dubai, United Arab Emirates

reteaching and revised support, not an accusation or automatic prohibition. Safety failures can require immediate stopping. These are proposed decision rules to test, not outcomes observed in a new pilot.

This yields a testable prediction: help is more likely to improve independent performance when it preserves the cognitive work needed for the target skill. Consultation logs alone cannot test that prediction. Researchers need evidence of what the learner attempted, what was delegated, how an answer was checked, and whether support was successfully withdrawn. Causal designs should examine these intermediates rather than equate interaction frequency with self-regulation.

Table 4: Proposed age–tool–supervision–unaided-learning decisions

Context / age	Tool and supervision	Unaided evidence / decision
Any age: access ineligible / unsafe	Do not bypass restrictions; use approved adult-mediated or non-AI alternatives	Access, safety, and accessibility are gates, not outcomes; do not proceed while unresolved
Primary, ≈6–11	Approved bounded ITS / teacher demonstrations; adult reviews help after an attempt	Explain a related item without help; fade hints; reteach if the targeted skill is still outsourced
Lower secondary, ≈12–14	Eligible bounded ITS or structured GenAI; teacher reviews outputs and verification	Delayed aligned task without AI; score separately from assisted work; pause or redesign if learning declines
Upper secondary, ≈15–18	Eligible disciplinary GenAI / ITS; document assistance, sources, revisions, and reasoning	Unaided transfer / oral defence plus local comparison; scale only with acceptable learning, equity, and safety evidence

Unvalidated proposal, not universal age rules. Product terms, law, readiness, accessibility, and risk can supersede it. Sources: this review, UNESCO/UNICEF, learning theory, and Kovari (2026).

3.7 Local implementation and evaluation

Figure 5 retains familiar organizational steps without presenting them as the novel contribution. Start with a defined learning problem, approved data/access conditions, prepared teachers, and a bounded pilot. Set baseline, comparison, incident, workload, and learning indicators before use. Assign responsibility for authorization and redress; offer equivalent alternatives. Scaling requires acceptable independent-learning, equity, and safety evidence, not popularity.

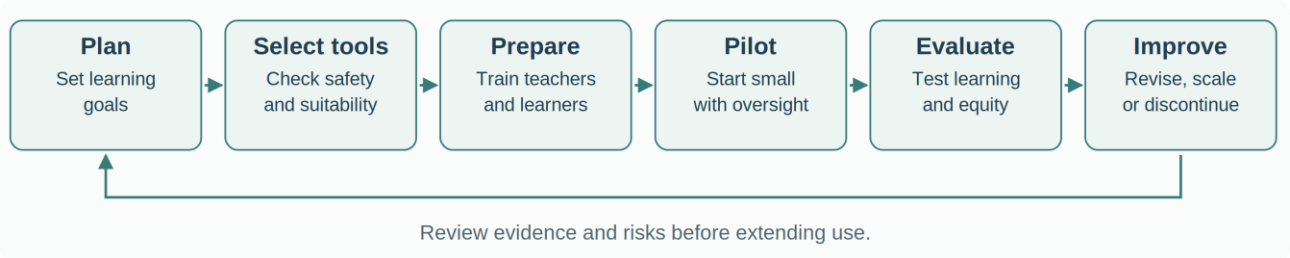


Figure 5: Plan, select, prepare, pilot, evaluate, and improve. Proposed implementation cycle, not validated

Measure immediate assisted output separately from delayed no-AI tasks and transfer; record reasoning, participation, accessibility, and net teacher time. A credible comparison can address teacher attention or novelty, but ordinary monitoring cannot establish causality. University evidence cannot establish children's suitability (Kestin *et al.,*2025; Zawacki-Richter *et al.,*2019), and early reviews had limited school evidence (Lo, 2023). Longer, more diverse school studies should test the proposed mechanisms and decisions.

4. CONCLUSION

AI can support school practice, access, and teacher preparation, but neither fluent output nor adoption establishes learning. The first question is answered conditionally: bounded tutoring and general-purpose GenAI are different technologies, and effects depend on comparison, design, duration, and assessment regime. The second concerns interconnected educational and ethical risks, including inaccurate content, inequitable access, privacy loss, and substituted practice. The third requires decisions tied to eligible tools, accountable supervision, and learning demonstrated when assistance is removed.

This synthesis contributes a transparent comparative evidence record and an unvalidated school-decision layer connecting age, tool type, supervision, and unaided-learning evidence. It complements, rather than replaces, existing competency, child-rights, and assessment frameworks. Self-regulation, offloading, and scaffolding make the proposed routine theoretically interpretable, but its mechanisms and effectiveness have not been established. The targeted search, incomplete historical screening records, English-language retrieval,

heterogeneous outcomes, and context-specific trials constrain conclusions. Nonsignificant results are not equivalence, and immediate tests are not delayed transfer. Future studies should verify intervention components, retain uncertainty, measure independent learning over time, include younger and diverse learners, and investigate equity, well-being, and teacher work. Educational goals and children's developmental needs should determine when AI is used, revised, or deliberately set aside. Any scaling decision must remain provisional and open to reversal.

REFERENCES

- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences of the United States of America, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122
- Fütterer, T., Bardach, L., Kuhn, J., Keller, S. D., & Gerjets, P. (2026). Enhancing school students' self-regulated learning through generative AI support: A randomized controlled trial. *Educational Psychology Review, 38*, Article 42. https://doi.org/10.1007/s10648-026-10133-8
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (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
- Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*, Article 17458. https://doi.org/10.1038/s41598-025-97652-6
- Kovari, A. (2026). Generative AI, performance, and learning: A framework for comparing interventions across assessment regimes. *Computers, 15*(9), Article 633. https://doi.org/10.3390/computers15090633
- 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
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), Article 410. https://doi.org/10.3390/educsci13040410
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727
- Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K–12 education. *npj Science of Learning, 10*, Article 29. https://doi.org/10.1038/s41539-025-00320-7
- 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
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104_eng
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693_eng
- Miao, F., Shiohira, K., & Lao, N. (2024). *AI competency framework for students*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391105_eng
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041
- Paglialunga, A., & Melogno, S. (2025). The effectiveness of artificial intelligence-based interventions for students with learning disabilities: A systematic review. *Brain Sciences, 15*(8), Article 806. https://doi.org/10.3390/brainsci15080806
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, Article 422. https://doi.org/10.3389/fpsyg.2017.00422
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
- Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open, 2*(4), Article 2332858416673968. https://doi.org/10.1177/2332858416673968
- Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970–987. https://doi.org/10.1037/a0032447
- UNESCO. (2023, September 7). *UNESCO: Governments must quickly regulate generative AI in schools*. https://www.unesco.org/en/articles/unesco-governments-must-quickly-regulate-generative-ai-schools

- UNICEF. (2025). *Guidance on AI and children: Version 3.0—Recommendations for AI policies and systems that uphold child rights*. UNICEF Innocenti. https://www.unicef.org/innocenti/media/11991/file/UNICEF-Innocenti-Guidance-on-AI-and-Children-3-2025.pdf
- 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
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x
- Yang, T.-C., Hsu, Y.-C., & Wu, J.-Y. (2025). The effectiveness of ChatGPT in assisting high school students in programming learning: Evidence from a quasi-experimental research. *Interactive Learning Environments, 33*(6), 3726–3743. https://doi.org/10.1080/10494820.2025.2450659
- 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
- AI assistance disclosure: Manus generative AI assistance supported literature locating, structured extraction, drafting, and language editing. Figures 1 and 3 are AI-generated conceptual illustrations; Figures 2, 4, and 5 are original vector syntheses. No participant data or experimental results were generated.