Artificial Intelligence in K–12 Schools

Cristina Barnard, Lily Fesler, and Susanna Loeb

Introduction

Schools are adopting artificial intelligence (AI) tools faster than researchers can evaluate them. Survey evidence suggests that student use of AI for school is already widespread. In a nationally representative 2025 RAND survey, 54% of middle and high school students reported using AI for school, and 53% of English language arts (ELA), math, and science teachers reported using AI in instructional tasks.[1] New systems promise to provide tutoring, generate instructional materials, deliver automated feedback, assist teachers with planning and assessment, and help schools coordinate learning opportunities.

Despite rapid growth in AI-related education research, the causal evidence base remains small, especially for U.S. K–12 settings. A recent review identified more than 800 relevant papers, but only 20 high-quality causal studies on impacts for students or educators.[2] The review found no high-quality causal studies conducted in U.S. K–12 schools on student-facing AI tools. Most of the studies examine short-term task performance rather than durable impacts on students' knowledge and skills.

This evidence gap creates a practical problem. Because new tools are emerging continuously and their capabilities are changing quickly, educators cannot rely only on product-specific studies to guide decisions. A tool evaluated rigorously today may look different by the time findings are published. Schools, therefore, need a framework for evaluating AI that can keep pace with technological change while remaining grounded in what research already shows about learning.

Research consistently demonstrates that cognitive, social, and emotional competencies influence both academic performance and long-term success.[3] Drawing from five established social and emotional learning (SEL) frameworks informed by research across education, psychology, and labor economics, we identify the following competencies for thriving: academic knowledge and skills, higher-order thinking skills, social skills, metacognition, self-regulation, adaptability, autonomy, motivation, interest, curiosity, belonging, interpersonal connection, self-efficacy, a growth-oriented mindset and self-concept, and management of content-specific anxiety, boredom, and frustration.[4]

The empirical evidence on which experiences develop these competencies is well established and independent of the rapid cycles in which technological solutions develop. Key among these experiences are personalized instruction, complex problem-solving, meaningful collaborative projects, substantive discussion, and sustained relationships with educators.[5] This chapter argues that the most useful way to evaluate AI in education is to ask whether AI tools expand students' access to those experiences. This question applies across technologies and connects adoption decisions to the educational goals schools are trying to achieve.

The chapter first reviews the evidence base and its limitations. It then examines student-facing AI tools, for which the latest causal evidence is clearest. Subsequent sections discuss design choices, educator-facing tools, system-level applications, and assessment. The chapter concludes with implications for practice and priorities for future research. The extensive list of priorities in the final section reflects the scarcity of empirical evidence on AI applications in K–12 settings.

Key findings

AI tools are entering schools quickly, yet rigorous evidence on their educational effects remains sparse. Early causal studies nonetheless suggest several consistent patterns.

Key finding #1: *The evidence base on AI in K–12 education—especially causal evidence on U.S. schools—is thin but growing.*

The related studies are often conducted in postsecondary or international settings, examine short-term uses of AI, and rely on narrow outcome measures.

Key finding #2: *AI tools influence learning through the experiences they create for students*.

Evidence on learning suggests that certain experiences are central to durable learning. These include personalized instruction, complex problem-solving, meaningful collaboration, substantive discussion, and sustained relationships with educators. AI tools potentially shape these experiences through student-facing interactions, educator support, and system-level organization of learning opportunities.

Key finding #3: *AI tools often improve performance during use, but these gains do not consistently reflect independent proficiency.*

Students using general-purpose AI systems frequently complete tasks more successfully when assisted, but may show weaker performance on unassisted assessments than students who did not have access to AI. In some cases, AI support may reduce opportunities for students to develop the skills needed to perform tasks independently. Tools that guide students through reasoning rather than generating answers appear likelier to support durable learning.

Key finding #4: *Emerging evidence suggests that design choices shape whether AI tools support or undermine learning.*

Systems that prompt students to explain their reasoning, engage actively with problems, and revise their understanding tend to reinforce the cognitive work associated with deeper learning. Systems that generate complete answers tend to reduce cognitive work and learning.

Key finding #5: *AI tools sometimes expand educator capacity, but their value likely depends on how freed-up time and attention are redirected..*

Early evidence suggests that generative AI can reduce the time educators spend on routine preparation and grading tasks without reducing the quality of the materials produced. Tools that provide instructional feedback to teachers and tutors also show early promise.

Evidence

Key finding #1: *The evidence base on AI in K–12 education—especially causal evidence on U.S. schools—is thin but growing.*

Research on AI in education is growing quickly, but rigorous evidence on its effects remains limited. A recent review identified more than 800 academic papers relevant to AI in K–12 education.[6] Only 20 of the studies meet the standards for strong causal inference, using randomized controlled trials (RCTs) or quasi-experimental designs to estimate impacts on students or educators.

According to this review, the high-quality causal evidence is especially thin for U.S. K–12 classrooms. The student-facing studies that met the review standards were conducted either in postsecondary settings or in K–12 settings outside the United States. The educator-facing evidence is somewhat more directly relevant, with several such studies conducted in U.S. K–12 settings.

The studies also examine a narrow range of uses and outcomes. Most of the student-facing studies focus on short-term interactions with AI tools, such as support on math problems, writing tasks, programming assignments, or brief learning activities. Many of the studies measure performance while students have access to AI or shortly after use, instead of tracking whether students develop durable knowledge and skills that transfer to independent work. The causal evidence is also concentrated in a small set of subjects, especially math and computer science, with much less evidence on literacy, social studies, language learning, collaboration, social-emotional development, or longer-term academic outcomes.

The current evidence provides early signals about student performance, teacher time use, and instructional support. It also highlights risks, including tool-dependent performance, reduced cognitive effort, and variation across students. The evidence does not yet show how AI tools affect learning across diverse classrooms, school systems, student populations, and implementation conditions. The broader research base includes descriptive and technical studies that clarify how tools are designed and used, but these studies do not establish the causal impacts of such tools on learning. More research is needed on longer-term outcomes, contextual variation, and implementation in U.S. K–12 classrooms.

Key finding #2: *AI tools influence learning through the experiences they create for students.*

Decades of research in the learning sciences establishes that durable learning requires students to actively engage with ideas, explain their reasoning, and revise their understanding in response to feedback.[7] Building on these principles, experimental studies have identified specific learning experiences that generate these conditions, including personalized instruction, complex problem-solving, substantive discussion, and sustained relationships with educators.[8] This evidence base points to a central evaluative question for AI tools: Does the tool expand students' access to the experiences that research identifies as important for development?

AI tools enter schools through three distinct pathways. Some interact directly with students by providing tutoring, feedback, or problem-solving support. Others support teachers and tutors by generating instructional materials, analyzing student work, or offering guidance on practice. Still others shape how schools organize learning by supporting scheduling, communication, or coordination across programs.

Each pathway matters because it influences the learning experiences students encounter. Student-facing tools shape how deeply students engage with academic tasks. Educator-facing tools affect how teachers and tutors allocate their time and attention. System-level tools determine which students gain access to which learning opportunities. Across these pathways, AI functions less as a standalone instructional intervention and more as a mediator of the learning experiences students encounter. The sections that follow examine each pathway in turn, beginning with the student-facing tools for which the latest causal evidence is strongest.

Key finding #3: *AI tools often improve performance during use, but these gains do not consistently reflect independent proficiency.*

A growing body of research examines AI tools that interact directly with students during learning activities. The clearest pattern in this literature is that students often perform better while they have access to AI support but these gains do not always carry over to independent work.

One well-designed example comes from a randomized study of high school students using generative AI for mathematics practice.[9] Compared with students who practiced without any AI support, students using a tutoring-style version of the system, which offered hints and guided reasoning rather than complete solutions, showed gains on practice tasks and no difference in performance on a subsequent closed-book exam. Students who had used a general purpose chatbot performed worse than students who had practiced with no AI support on the subsequent closed-book exam, even though they had higher performance on practice tasks.

Other studies document similar patterns across subjects and age groups. Experiments with AI tutoring systems and automated feedback tools show short-term improvements in homework performance, writing revision, and problem-solving during assisted activities.[10] The evidence on independent learning is more mixed. A study that assessed student performance without AI assistance found weaker or no evidence of learning gains.[11] Students using ChatGPT for a scientific inquiry task reported lower cognitive effort and produced weaker reasoning and argumentation than students using traditional search engines.[12]

Some evidence points to more promising uses of AI-supported tutoring when the tools are implemented under appropriate enabling conditions. A quasi-experimental investigation of hybrid human–AI tutoring systems found improvements in student learning across three studies, suggesting the pattern holds across different implementations.[13] Another experimental study found that students with access to an AI tutor learned more than students using textbooks alone, although students with unrestricted AI access spent less time reading textbook materials within a fixed learning period.[14] These findings suggest that AI tools may support learning when they bolster rather than replace student effort.

These studies point to a meaningful distinction between task performance and learning. AI systems can help students complete work successfully in the moment without building the knowledge and skills they need to perform independently later. The evidence also suggests a second concern: Some tools may reduce the cognitive effort students need to develop independent proficiency. To understand why some tools produce more durable gains than others, we must look more closely at how these systems are designed.

Key finding #4: *Emerging evidence suggests that design choices shape whether AI tools support or undermine learning.*

The mixed evidence on AI tools that interact directly with students has a consistent explanation: The technology itself matters less than how it structures students' engagement with learning tasks. AI systems that generate complete solutions can short-circuit the processes students need to go through to learn. When students receive answers immediately, they may complete assignments successfully without performing the cognitive work that research identifies as crucial for building lasting knowledge.[15]

Design choices appear central to whether AI tools expand or reduce students' access to the kinds of cognitive engagement that support learning. Tools that prompt explanation, reflection, and revision strengthen the learning experiences research identifies as most important. Tools that generate answers efficiently can undermine those same experiences even when they appear to improve performance in the moment.

Experimental evidence supports this mechanism for short-term learning gains. Experiments explicitly comparing outcomes between students who engaged with AI tools that generate step-by-step explanations and students who used tools with direct answer generation have found that guided reasoning improves students' problem-solving performance on assessments completed without AI support.[16] Similarly, one experimental study found that students using a tutoring-style AI system that emphasized hints and guided reasoning performed well during practice without the declines in subsequent closed-book exam performance observed among students using a general-purpose chatbot.[17] However, on closed-book exams, general purpose chatbot users performed worse despite their comparable practice performance, while tutoring-AI users performed no better than students without AI support. More research could clarify the optimal design features and conditions for promoting lasting learning.

Key finding #5: *AI tools sometimes expand educator capacity, but their value likely depends on how freed-up time and attention are redirected.*

AI tools may influence learning indirectly by changing how educators allocate their time and attention across instructional activities. When we consider tools that allow educators to save time without compromising quality, the central question becomes whether any time saved on routine tasks is redirected toward the kinds of interactions—individualized instruction, discussion, facilitation of group activities, and mentorship—that research identifies as most valuable for student development.

The early evidence on time savings is encouraging. In a randomized experiment, teachers with access to ChatGPT spent less time preparing lessons while producing instructional materials that external reviewers rated as comparable in quality to those produced by teachers without AI access.[18] Research on teacher task allocation suggests that AI tools may reduce time spent on routine activities while creating space for more individualized instruction.[19]

Other tools go beyond time savings by providing direct guidance on instructional practice. Evidence suggests that automated systems that analyze classroom dialogue help teachers reflect on their questioning strategies and improve how they respond to student ideas during instruction.[20] Earlier work on AI-supported classroom analytics found that systems designed to increase teachers' awareness of student activity can improve instructional decision-making in technology-enhanced classrooms.[21] Similarly, a randomized field experiment of technology-assisted home tutoring found that diagnostic reports helped tutors identify student progress and areas of difficulty, improving student academic performance.[22] In a randomized study of remote tutoring, an AI system that generated instructional suggestions for the tutor improved student learning outcomes, with the largest gains for tutors with lower prior performance ratings.[23]

Tools that provide instructional guidance are particularly promising because they expand what educators can do with the time they have. Whether educator-facing AI ultimately improves student learning depends on whether schools create the conditions for that redirection to happen.

Areas for future research

Student learning outcomes

Most studies of AI tools aimed at improving student learning explore short-term outcomes,[24] such as achievement scores measured right after the practice sessions with the AI tools. The long-term effects of these tools on student achievement and on broader competencies remain largely unexplored. Further research should examine the effects of AI tools on end-of-year academic achievement over extended implementation periods and on medium-term outcomes for competencies such as motivation, metacognition, critical thinking, and self-regulation. This research should be conducted under real-world conditions in school, and it should include measures of implementation quality. Given evidence from other educational interventions that short-term gains often fade over time, longer-term follow-up may be especially important for AI tools that improve task performance during use without strengthening students' independent proficiency.

An essential question is whether AI tools function as effective scaffolds for durable learning. Accordingly, studies should include achievement outcomes measured in the absence of the tool to assess whether students can transfer their learning to independent performance.

Equity and student experiences

AI has the potential to expand access to learning experiences that have historically been unevenly distributed, including personalized instruction, complex problem-solving, meaningful collaboration, and sustained relationships with educators. Differences in access to high-quality tools and implementation capacity may also introduce new inequalities.

The distribution of learning experiences remains uneven across students and schools. While tools that expand access to these experiences may reduce disparities, uneven access to effective tools or implementation supports may widen them. The effects of AI adoption will depend on how tools are implemented across contexts.

AI is also beginning to shape students' social and emotional experiences. Some students report using conversational AI systems as social companions.[25] The implications of such uses for well-being, identity development, and relationships with peers and educators remain unclear and require further study.

System-level applications

Many proposed applications of AI involve changes to how schools organize learning opportunities. These applications include tools designed to support scheduling, coordinate instructional services, improve communication with families, recommend flexible student groupings, and allocate tutoring or mentoring resources across students. Rigorous causal evidence on these applications remains sparse. Most of the related studies focus on tools that interact directly with students or support individual educators, leaving system-level applications largely unevaluated.

This evidence gap is worth taking seriously because organizational constraints often limit schools' ability to provide high-quality learning experiences. Schools may struggle to deliver tutoring, mentorship, or project-based learning at scale because of scheduling, staffing, and coordination challenges. AI systems that support scheduling, match students with tutors or mentors, coordinate apprenticeships, or organize small-group instruction could expand access to these opportunities.

Consider a middle school that aims to organize mornings around literacy and mathematics instruction and afternoons around interdisciplinary projects. Implementing this schedule requires coordination across staffing, classroom space, student progress, and intervention needs. Systems that integrate learning data with staffing and scheduling constraints could support these designs by dynamically recommending student groupings, adjusting instructional time, and identifying students who need additional support.

The framework developed in this chapter applies directly to such tools. The relevant question is whether AI applications help schools expand access to personalized instruction, complex problem-solving, meaningful collaboration, substantive discussion, and sustained relationships with educators. Systems that support these experiences could have meaningful effects on student learning. Systems that focus only on efficiency may reinforce constraints.

Assessment

AI may also reshape how schools assess student learning. Assessment systems rely heavily on formats that are feasible to administer and score at scale, including short written responses and selected-response questions. These formats often emphasize recall and procedural knowledge, even when schools aim to develop reasoning, communication, and complex problem-solving.

AI tools can evaluate extended writing, mathematical reasoning, and problem-solving processes. These capabilities create opportunities to assess how students develop ideas, revise work, and contribute to collaborative tasks. These forms of assessment align more closely with the learning experiences that support student development.

The early evidence on AI-generated feedback is promising but limited. Studies find that AI-generated feedback can increase revision rates and student engagement, although human feedback remains of higher quality overall.[26] These findings suggest that AI may expand access to formative feedback, particularly in early stages of work.

AI-based assessment also raises challenges. Automated scoring systems may vary in reliability across tasks and student populations. Concerns about bias remain, especially when systems rely on historical data. Students may also adapt responses to optimize scores without demonstrating deeper understanding. Questions of validity remain central, particularly when assessments rely on process data such as revision histories or collaboration patterns.

Research Priorities

We highlight several priorities for future research. Studies that track durable learning outcomes are needed to clarify whether AI tools support the development of knowledge and skills over time. Research on implementation across diverse school contexts can clarify how tools function under real conditions. In addition, work that examines variation across student populations is needed to identify which students benefit most and under what conditions.

Research that connects system-level implementation with student outcomes will be especially valuable. Collaboration between researchers, educators, and developers can support the design and evaluation of tools that expand access to effective learning experiences.

AI is likely to remain a central part of K–12 education. Its educational significance will depend on how these tools are integrated into teaching and learning and whether they expand access to the experiences that support student development.

Endnotes and references

---

[1] Doan, S., E. D. Steiner, and A. Woo. 2025. AI Use in Schools Is Quickly Increasing, But Guidance Lags Behind: Findings from the RAND Survey Panels. RAND Corporation.
[2] Fesler, L., A. Martinez Claeys, W. Agnew, and S. Loeb. 2026. *Artificial Intelligence in K–12 education: A Review of the Evidence*. Stanford University, SCALE Initiative.
[3] Durlak, J. A., C. E. Domitrovich, R. P. Weissberg, and T. P. Gullotta (Eds.). 2015. *Handbook of Social and Emotional Learning: Research and Practice* (pp. xxii, 634). The Guilford Press; Jones, S. M., and E. J. Doolittle. 2017. Social and Emotional Learning: Introducing the Issue. *The Future of Children*, 27(1), 3–11.
[4] Collaborative for Academic, Social, and Emotional Learning. 2020. CASEL's SEL Framework: What Are the Core Competence Areas and Where Are They Promoted? October 1; Farrington, C. A., M. Roderick, E. Allensworth, J. Nagaoka, T. S. Keyes, D. W. Johnson, and N. O. Beechum. 2012. *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance–A Critical Literature Review*. Consortium on Chicago School Research; Organisation for Economic Co-operation and Development. 2019. OECD Learning Compass 2030. *OECD Future of Education and Skills 2030*; Partnership for 21st Century Skills. 2009. P21 Framework Definitions; Ryan, R. M., and E. L. Deci. 2000. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist* 55(1), 68–78.
[5] Barnard Gonzalez, C., C. Agnew, and S. Loeb. 2026. *The Learning Experiences That Matter and AI's Role. Getting Down to Facts III*. Stanford University, SCALE Initiative.
[6] Fesler et al., 2026.
[7] National Academies of Sciences, Engineering, and Medicine. 2018. *How People Learn II: Learners, Contexts, and Cultures.* National Academies Press.
[8] Barnard Gonzalez et al., 2026.
[9] Bastani, H., O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman. 2025. Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics. *Proceedings of the National Academy of Sciences* 122(26), e2422633122.
[10] Chen, E., S. Judicke, K. Beigh, X. Tang, Z. Xiao, C. Li, S. Li, R. Luttmer, S. Singh, M. Yampolsky, N. Parikh, Y. Zhao, M. Chen, S. Huang, A. Mohanty, G. Johnson, J. Mackey, J. Lin, and K. Koedinger. 2025. Generative AI Alone May Not Be Enough: Evaluating AI Support for Learning Mathematical Proof [Preprint]. arXiv; Meyer, J., T. Jansen, R. Schiller, L. W. Liebenow, M. Steinbach, A. Horbach, and J. Fleckenstein. 2024. Using Large Language Models to Bring Evidence-Based Feedback into the Classroom: AI-Generated Feedback Increases Secondary Students' Text Revision, Motivation, and Positive Emotions. *Computers and Education: Artificial Intelligence* 6, 100199; Zhao, C., M. Silva, and S. Poulsen. 2025. Autograding Mathematical Induction Proofs with Natural Language Processing [Preprint]. arXiv.
[11] Chen et al., 2025.
[12] Stadler, M., M. Bannert, and M. Sailer. 2024. Cognitive Ease at a Cost: Large Language Models Reduce Mental Effort but Compromise Depth in Student Scientific Inquiry. *Computers in Human Behavior* 160, 108386.
[13] Thomas, D. R., J. Lin, E. Gatz, A. Gurung, S. Gupta, K. Norberg, S. E. Fancsali, V. Aleven, L. Branstetter, E. Brunskill, and K. R. Koedinger. 2024. Improving Student Learning with Hybrid Human–AI Tutoring: A Three-Study Quasi-Experimental Investigation. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 404–415). ACM.
[14] Fischer, M., H. A. Rau, and R. M. Rilke. 2025. AI Tutoring Enhances Student Learning Without Crowding Out Reading Effort. IZA Discussion Paper No. 18338.
[15] National Academies of Sciences, Engineering, and Medicine. 2018. *How People Learn II: Learners, Contexts, and Cultures*. National Academies Press.
[16] Blasco, A., and V. Charisi. 2025. AI Chatbots in K–12 Education: An Experimental Study of Socratic vs. Non-Socratic AI Agents and the Role of Step-by-Step Reasoning. SSRN Electronic Journal.
[17] Bastani, H., O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman. 2025. "Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics." *Proceedings of the National Academy of Sciences* 122(26): e2422633122.
[18] Roy, P., H. Poet, R. Staunton, K. Aston, and D. Thomas. 2024. *ChatGPT in Lesson Preparation: A Teacher Choices Trial.* National Foundation for Educational Research.
[19] Ferman, B., L. Lima, and F. Riva. 2021. Artificial Intelligence, Teacher Tasks and Individualized Pedagogy. J-PAL Working Paper.

[20] Demszky, D., J. Liu, H. C. Hill, D. Jurafsky, and C. Piech. 2023. Can Automated Feedback Improve Teachers' Uptake of Student Ideas? Evidence from a Randomized Controlled Trial in a Large-Scale Online Course. *Educational Evaluation and Policy Analysis*; Demszky, D., J. Liu, H. C. Hill, S. Sanghi, and A. Chung. 2025. Automated Feedback Improves Teachers' Questioning Quality in Brick-and-Mortar Classrooms: Opportunities for Further Enhancement. *Computers & Education* 227, 105183.

[21] Holstein, K., B. M. McLaren, and V. Aleven. 2018. Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms. In *Artificial Intelligence in Education*. Edited by C. Penstein Rosé et al. (Lecture Notes in Computer Science, Vol. 10947, pp. 154–168). Springer.

[22] Kim, J. H., M. Kim, D. W. Kwak, and S. Lee. 2022. Home-Tutoring Services Assisted with Technology: Investigating the Role of Artificial Intelligence Using a Randomized Field Experiment. *Journal of Marketing Research* 59(1), 79–96.

[23] Wang, R. E., A. T. Ribeiro, C. D. Robinson, S. Loeb, and D. Demszky. 2025. Tutor CoPilot: A Human–AI Approach for Scaling Real-Time Expertise [Preprint]. arXiv.

[24] Fesler et al., 2026.

[25] Robb, M. B., and S. Mann. 2025. *Talk, Trust, and Trade-offs: How and Why Teens Use AI Companions.* Common Sense Media.

[26] Meyer et al., 2024; Steiss, J., T. Tate, S. Graham, J. Cruz, M. Hebert, J. Wang, Y. Moon, W. Tseng, M. Warschauer, and C. B. Olson. 2024. Comparing the Quality of Human and ChatGPT Feedback of Students' Writing. *Learning and Instruction 91*, 101894.