AI and Learning in K-12 Education

What Recent Research Shows — and What It Means for AI Policy

K. Sudhir
James L. Frank Professor of Marketing, Private Enterprise, and Management · Yale School of Management
Adapted from research prepared for a panel discussion at the Yale Education Leadership Conference, April 24, 2026

PURPOSE

This brief is intended to help district and school leaders think through AI policy with the benefit of current research. It does three things: it summarizes what recent studies show about when AI helps learning and when it harms it; it draws one policy implication that follows from reading the findings together; and it offers questions and vocabulary that leadership teams can use in their own discussions.

Well-designed AI can produce substantial learning gains; poorly designed AI can harm learning in ways students do not recognize. The findings below document both sides of that range and the design choices that determine which side a deployment falls on.

A note on the evidence: This is an early and developing research area. The studies summarized here are from 2025–2026; they are short-term, conducted in specific subject areas, and not yet replicated at scale. The findings are directional rather than settled. A different researcher would select a somewhat different set of studies; this brief privileges causally identified work — randomized trials and quasi-experimental research — for findings on learning outcomes, and descriptive studies for findings on AI grading.

WHAT THE RESEARCH SHOWS

Finding 1 — Whether AI helps or harms learning depends on how it is designed, not which model is used

A Wharton study (Bastani et al., 2025) randomly assigned roughly 1,000 high school math students to three conditions: no AI, AI with no guardrails, or the same AI redesigned to give hints rather than answers. Students with unguarded AI practiced 48% more than the control group — and scored 17 percentage points lower on a subsequent exam taken without AI access. Students with guardrailed AI practiced 127% more and showed no exam harm. The difference was the interaction design. A Harvard randomized trial on physics tutoring (Kestin et al., 2025, N=194) found that pedagogically designed AI tutoring produced two times the learning gains of the best available active learning classroom instruction, in less time. A Harvard and Penn study (Lira, Rogers & Duckworth, 2026) found AI writing coaching produced genuine skill improvement — an effect size of 0.40 — exceeding the gains from professional human editor feedback. In the available evidence, the ceiling for well-designed AI in education appears high and the floor for poorly designed AI appears harmful in ways invisible to students themselves. The high-end results come from tightly designed environments and may not generalize automatically to typical classroom deployments.

Finding 2 — Engagement metrics are an unreliable measure of whether AI is producing learning

Preliminary research on a Japanese online learning platform (Zhu, Sudhir, and Uetake — Yale Working Paper, 2026) examined how AI-mediated goal structure affects both completion rates and quiz accuracy. Students given small, frequent goals showed high completion rates but no improvement in quiz scores. Students given medium-sized goals showed both higher completion and genuine learning gains. This pattern — high engagement metrics alongside flat learning outcomes — is consistent with the Bastani study, where students using unguarded AI reported feeling they had learned more than those in the guardrailed condition, despite scoring significantly worse on independent assessment. MIT neuroscience research (Kosmyna et al., 2025) adds a mechanistic signal: in the experimental setting, AI-assisted essay writing reduced neural engagement by 55%, and most participants could not accurately recall the content of essays they had just written with AI help. These findings suggest that completion rates, login counts, and student satisfaction scores are insufficient measures of whether an AI tool is producing learning. Independent assessment of learning outcomes is necessary.

Finding 3 — AI grading has a measurable performance profile and equity risks that warrant careful policy attention

On structured tasks with clear rubrics, AI grading can match human feedback while dramatically improving speed and consistency. A meta-analysis of 41 studies covering nearly 5,000 students found that AI feedback produces equivalent learning gains to human feedback when delivered promptly. AI grading shows consistent accuracy limitations on open-ended tasks, however, with human marker agreement of r=0.60–0.75 compared to r=0.85–0.95 on structured tasks. Studies have also documented demographic bias: AI graders scored Black students lower than Asian students even after controlling for essay quality; assigned lower scores to essays when informed the student attended an inner-city school; and systematically underscored English Language Learner students (Warr & Heath, 2025; Guo et al., 2025). An important methodological caveat: these studies used zero-shot prompting, without human-graded exemplars to anchor the AI's judgments. Rubric-anchored prompting with exemplars may reduce some of the racial disparities and grade compression effects, though the ELL and socioeconomic findings are less likely to be explained by prompting alone. The appropriate response is not to avoid AI grading, but to require vendors to demonstrate bias testing under rubric-anchored deployment conditions on disaggregated populations, and to keep human review on consequential grades.

RELATED EVIDENCE FROM ADJACENT SETTINGS

One pattern from adjacent research bears mention even though it has not yet been directly tested in K-12 settings. A study of consultants using GPT-4 (Dell'Acqua et al., 2023) found that while individual performance on the assigned tasks improved, the range and distinctiveness of ideas generated across users narrowed. Different consultants, given the same problems, produced more similar outputs when working with AI than they did without it.

The K-12 application is an extrapolation, but a directionally relevant one. If students routinely route their thinking through the same model, their analytical framings, examples, and arguments may converge in similar ways. This is a concern not only for individual student creativity but for the diversity of perspective that education is meant to cultivate. The K-12 evidence on this is not yet available; the consultant finding is offered as a directionally relevant signal worth monitoring, not a settled K-12 result.

AN IMPLICATION FOR POLICY

The three findings above come from individual studies. The following is an inference drawn from reading them together — a policy implication rather than a direct research result. It is offered as a frame for discussion and planning, not as a conclusion from any single study.

Appropriate AI use may differ meaningfully by course type and student developmental stage

A single, uniform AI policy across all subjects and grade levels is probably too blunt. The research on when AI harms learning tends to point to situations where AI substitutes for the cognitive work that builds foundational skills. The research on when AI helps tends to point to situations where foundations exist and AI extends what the student can do rather than bypassing what the student needs to do. Courses where core cognitive capacities are still forming — foundational math, introductory writing, core science — may warrant a different approach than project-based work, advanced courses, and electives where foundations are established. This applies throughout K-12: science fair projects, research papers, capstone assignments, and independent study all fit the second category. In foundational contexts, the key question is whether AI is configured so that students do the cognitive work. In project-based and advanced contexts, the relevant question shifts toward whether students can direct AI effectively, evaluate its outputs critically, catch errors AI cannot self-detect — and whether sustained AI use is producing skill atrophy in capacities students had already built. Many of these risks arise not from the AI itself, but from how systems are designed and what they are incentivized to optimize. This implication is consistent with the three research findings but is not itself a tested empirical result.

QUESTIONS FOR SCHOOL LEADERSHIP TEAMS

Discussion prompts for district and school leadership, AI policy committees, and curriculum teams — not recommendations.

Learning outcomes

1. When evaluating AI platforms for classroom or homework use, does the district have a way to assess learning outcomes — not just engagement metrics like completion rates and time on task?
2. Does the district's AI use policy distinguish between courses where foundational skills are still forming and contexts where those foundations are established — including project-based, capstone, and integrative assignments where students bring foundational skills to bear on open-ended problems?

Equity and accountability

3. For any AI grading or feedback tools in use or under consideration, has the district requested bias testing data from vendors — on race, socioeconomic context, and ELL status — under rubric-anchored deployment conditions?
4. What metrics are vendors optimizing for, and do these align with the district's definition of learning?

Teacher judgment

5. Does the curriculum include instruction in how to evaluate AI outputs critically — not just how to use AI tools — so that students develop the judgment that effective AI use requires?
6. What role do teachers play in decisions about which AI tools enter classrooms, and how is teacher judgment integrated into AI use policy?

EIGHT TERMS FOR POLICY DISCUSSIONS

Term	What it means
Engagement trap	AI optimizes for the feeling of learning rather than learning itself — producing high engagement metrics while learning outcomes remain flat or decline.
Scaffolded struggle	Assignments designed so that the productive difficulty of working through a problem is the pedagogical point. AI may be present but is configured so that the student does the cognitive work.
Feedback latency	The time between when a student submits work and when they receive feedback on it. Students act on feedback within hours but rarely on feedback arriving weeks later. AI can dramatically reduce feedback latency on structured work — one of the clearest opportunities in the research.
Practice bypass	Skills develop through the practice of doing the work, often imperfectly and repeatedly. AI makes it easy to bypass that practice — and the judgment and fluency that practice builds cannot form without it.
Cognitive offloading	Delegating thinking to an external tool. Moderate offloading is part of healthy cognition — calculators, reference texts, and search engines all involve offloading. The concern with AI is that it can offload the reasoning itself, not just the computation or lookup, and in ways that are invisible to students.
Cognitive foreclosure	Foundational cognitive capacities that were never built may be difficult to rebuild later — distinct from skill atrophy, which is recoverable. Applies wherever core skills are still forming. Results from sustained cognitive offloading during the formative period.
Skill atrophy	Capacities built in the past weaken from disuse when AI handles work that would have exercised them. Unlike foreclosure, atrophy is recoverable — but only if deliberate practice without AI is maintained.
Cognitive homogenization	When many students run their thinking through the same AI model, the diversity of perspective that education is supposed to cultivate may narrow toward the model's training-data center.

Disclosure: One of the studies cited — Zhu, Sudhir, and Uetake (Yale Working Paper, 2026) on goal structure and learning outcomes in an online learning platform — is preliminary work co-authored by me and has not yet been peer-reviewed. Other studies referenced are published or under review at peer-reviewed journals.

Suggested citation

Sudhir, K. (April 2026). AI and Learning in K-12 Education: What Recent Research Shows and What It Means for AI Policy. Yale School of Management research brief.

Citations for the studies referenced in this brief are available on request.