

Education AI's New Test Is What Students Can Do Without It

AI in EdTech Weekly

2026-08-24

Education AI's New Test Is What Students Can Do Without It

By AI in EdTech Weekly • August 24, 2026

This brief examines the shift from AI features to durable learning, as new federal edtech guidance arrives alongside direct-to-student rollouts, Alpha School's expansion, and growing pressure on educators and institutions to prove what AI actually improves.

Education AI is entering an evidence-and-design phase. The practical question is no longer whether a system can make a student's work look better, but whether the student can still perform when the system is removed. A current analysis of 27,000 students in China, new U.S. federal guidance, and a wave of student-facing product launches all point to that test. [1, 2]

The learning test is what remains when AI is gone

The clearest current signal is a short-term/long-term split. The Economist analysis reports that students using AI saw homework scores rise 18% and spent less time on assignments, but scored 20% worse on exams than students who did not use AI. The result does not make every AI use harmful; it does show why homework performance is becoming a weak proxy for learning when the tool does the hard cognitive work. [1]

Stanford's SCALE brief offers a useful filter for product claims. High-impact tutoring remains defined by live, human-led instruction; AI-led tools may provide supplemental practice, but they do not inherit that evidence base simply because they give one-to-one responses. The established tutoring literature reports average effects of 0.29–0.37 standard deviations, with stronger programs meeting frequently, connecting to classroom curriculum, adjusting from evidence of student understanding, maintaining a relationship over time, and training and supporting tutors. [3, 1]

A randomized field experiment with nearly 1,000 high-school mathematics students makes the distinction concrete. Unrestricted GPT-4 increased practice performance by 48%; a guardrailed “GPT Tutor,” supplied with teacher-written solutions and instructed to give hints, increased it by 127%. Once AI was removed, the general-chatbot group scored 17% below the control group on an unaided test, while the guardrailed group performed about the same as control. Guardrails prevented the measured harm, but did not automatically produce a learning gain. [1]

The evaluation problem is still visible at the model level: the best of 16 frontier models scored about 56/100 on TutorBench; almost no model-and-agent combination sustained strong tutoring across a simulated 30-day relationship; and six frontier models gave the answer in 97% of conversations with simulated eighth-graders when given generic instructions to be helpful. The last result is a diagnostic simulation, not a classroom failure rate, but it shows that “helpful” and “tutoring” are different objectives. [1]

Federal guidance meets AI-first school politics

The U.S. Department of Education’s new Dear Colleague Letter turns that learning question into a selection standard. It says technology should have a clear instructional purpose, be backed by independent evidence, be transparent about design and data use, and have strong implementation. States and districts are told to focus on learning impact, choose tools with proven results, review effectiveness regularly, and support educator training; repeated failure should lead to removal. The department’s five questions are practical: What problem does the tool solve? When and for whom should it be used? For how long? What evidence shows that it improves learning? [2]

The guidance favors instructional value over screen exposure, but leaves decisions to states and local communities. That creates a meaningful evidence bar without a single national implementation regime. EdSurge’s account notes that the department declined to offer new rules, while experts warned that privacy, accessibility, transparency, and responsible data use should not depend entirely on a student’s state. [2, 4]

The same week, federal education rhetoric gave an AI-led private-school model a prominent platform. Secretary Linda McMahon posted that responsible classroom AI can “revolutionize” learning and said she had seen it firsthand at Alpha School in Austin. Alpha is expanding from roughly a dozen to about 50 campuses; its model uses AI apps for core subjects such as math and reading for two hours, then spends the rest of the day on activities such as robotics and surfing. Co-founder Mackenzie Price says students learn twice as much in a fraction of the time, while adults focus on motivation, connection, and life skills. Those are the co-founder’s claims, not a comparison study presented in the interview. [5, 6]

That makes Alpha a useful test case, not a settled model. The department is ask-

ing for independent, condition-specific evidence, while SCALE classifies AI-only tutoring effectiveness as unknown and flags dosage and integration as unresolved



problems. [2, 3]

School uses AI instead of teachers for personalized learning: Co-founder breaks it down (0:00)

Student-facing AI is scaling before its evidence is settled

Google is distributing a study environment directly to eligible college students: one year of Google AI Pro at no cost in the United States and one year of Google AI Plus in 140-plus countries, with the offer running through December 31, 2026. Its student hub includes study notebooks, flashcards, and practice quizzes. Google says the notebooks can use diagnostic quizzes to identify knowledge gaps, generate bite-sized lessons grounded in class materials, and track progress on a live dashboard. [7, 8, 9]

The rollout also treats AI as a learning interface rather than only a text box: Gemini can generate interactive 3D simulations and dynamic tables, and Gemini Live can run a multi-step research report in the background for the student to discuss by voice later. [10, 11]

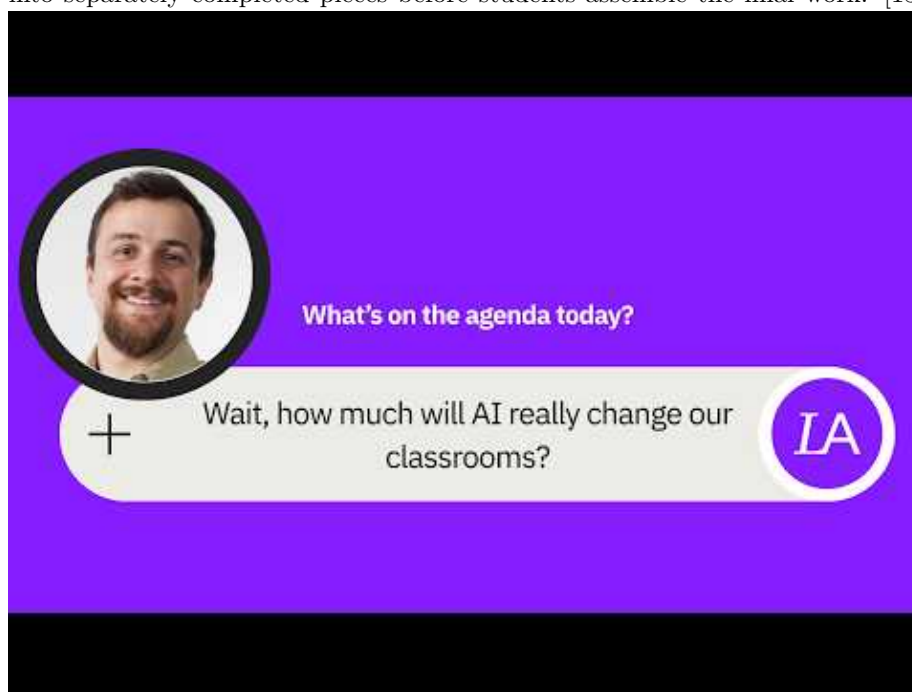
OpenAI is making a parallel move for younger learners. Its August 18 launch of ChatGPT for Teens targets ages 13–17 and automatically places users into the teen experience when the system estimates they are under 18. Study Mode, homework reminders, quizzes, Learning Visualizations, and Study Hours are designed to guide students through work instead of supplying a quick answer.

The product also adds default age-appropriate protections, Quiet Hours and selected parental controls, safety notifications in limited high-risk situations, break reminders, and restrictions against emotional dependence. [12]

The design direction is more disciplined than a generic answer engine, but the learning question remains open. The current tutoring analysis calls these structured products meaningful design shifts while explicitly warning that stated pedagogy is not yet evidence of durable learning. [1]

Teachers and leaders are absorbing the implementation cost

Justin Reich’s current interview, informed by conversations with 120 K–2 teachers and students, reaches a restrained conclusion: AI might modestly help some students learn some things, but is causing “a lot of headaches” and is not transformational. One immediate classroom response is defensive. Teachers are moving writing onto paper and into class to reduce chatbot completion, which protects authenticity at the cost of instructional time. Some are paying out of pocket for revision-history tools, yet students can still retype AI-generated work. Reich’s preferred “layers of resistance” approach breaks an essay into separately completed pieces before students assemble the final work. [13]



Why AI might not Change Education Much at All (with Justin Reich) (13:28)

The leadership capacity gap is just as concrete. In a June survey discussed by Vicki Davis and National Principals Association CEO Ronn Nozoe, seven in ten respondents said their schools were already immersed in AI, but only about

18% felt very confident leading implementation. Forty percent specifically asked for time for staff training and planning. The schools making progress, the speakers say, are doing so through administrator–teacher collaboration rather than a top-down mandate or an isolated “Lone Ranger” teacher. [14]



Principals and AI: 7 in 10 Are Already In (5:21)

Higher education is moving from chat to delegated work

An EDUCAUSE Exchange discussion describes agentic AI as systems that connect multiple steps, use tools, respond to events, and act across institutional workflows. Early use is most visible in research: the speaker estimates roughly 5% of researchers are using agentic tools, from Claude Code and Codex to networks of agents that search literature and pass summaries between one another. The governance issue is ownership: institutions that choose tools and write policies should carry substantial responsibility when agents act on behalf of staff or students. [15]

The human role is shifting toward judgment, but the discussion warns that a nominal human-in-the-loop can become “agency laundering,” with people approving an agent’s work automatically. It also frames “cognitive triage”—deciding what to delegate and what to retain—as a basic skill. Whether saved time improves advising and student support or simply becomes more cases and higher output is an institutional choice. [15]

For lifelong learning, the pressure is structural rather than cyclical. The World

Economic Forum expects 39% of workers’ core skills to change by 2030; U.S. job postings show 32% of required skills changing between 2021 and 2024, with one in four jobs experiencing 75% skill turnover. Higher education therefore needs recurring, shorter bursts of upskilling tied to different industry speeds and learner backgrounds, not only occasional returns for another full degree. [16]

What This Means

- **For K–12 leaders:** Turn the Department’s five questions into a procurement record for every AI tool: the learning problem, intended users, duration, implementation owner, and evidence of impact. Separate instructional use from recreational screen exposure, and make removal a real option when results do not hold. [2, 4]
- **For assessment leaders:** Add an unaided transfer check to any workflow in which AI improves the submitted artifact. Staged drafts, explanations, and observed performance can protect learning without making continuous surveillance the default; the current classroom response shows that anti-cheating measures can consume instructional time. [1, 13]
- **For edtech buyers and product teams:** Evaluate the tutor, not just the model. A credible student-facing tutor should pursue a defined objective, elicit evidence of understanding, adapt its next move, and keep the student responsible for the thinking; test engagement, dosage, and performance after AI is removed. [1, 3]
- **For higher-ed and L&D teams:** Assign clear ownership for agents and protect the time they save for judgment, feedback, and relationships rather than automatically increasing caseloads. Design recurring learning in short, rigorous units that match the pace of work and the learner’s existing expertise. [15, 16]

Watch This Space

- **Outcome evidence for direct-to-student AI:** Google’s student hub, OpenAI’s teen experience, and Alpha’s expansion all need measures beyond usage or polished work: sustained engagement, unaided performance, learning transfer, and the dosage required to produce it. [9, 12, 3]
- **Federal guidance becoming local practice:** Watch whether states and districts convert the new evidence language into performance-based contracts, independent evaluations, transparent data-use requirements, and actual removal of tools that fail. [4]
- **The reality of human oversight:** In higher education, the key test for agentic systems will be whether “human in the loop” remains attentive judgment rather than automatic approval—and whether saved time reduces burnout instead of triggering the academic ratchet. [15]

Sources

1. Homework Helper or AI Tutor? The Difference Determines What Students Learn.
2. U.S. Department of Education Releases Guidance on Responsible Use of Education Technology in the Classroom
3. AI Tutoring is Not a Monolith: What We Actually Know
4. Department of Education Issues Long-Awaited Edtech Guidance for States and Districts
5. X post by @EDSecMcMahon
6. School uses AI instead of teachers for personalized learning: Co-founder breaks it down
7. X post by @GeminiApp
8. X post by @GeminiApp
9. X post by @GeminiApp
10. X post by @GeminiApp
11. X post by @GeminiApp
12. Introducing ChatGPT for Teens: Built for learning, backed by protections | OpenAI
13. Why AI might not Change Education Much at All (with Justin Reich)
14. Principals and AI: 7 in 10 Are Already In
15. 4 Ways Agentic AI Will Reshape Higher Ed Work | EDUCAUSE Exchange
16. Faster Technology Cycles Mean Faster Retraining Cycles