

AI Tutoring's Next Test Is Whether Learning Survives the Tool

AI in EdTech Weekly

2026-08-31

AI Tutoring's Next Test Is Whether Learning Survives the Tool

By AI in EdTech Weekly • August 31, 2026

A teacher-led Gemini trial reports math gains while Tennessee evidence shows optional Khanmigo use is thin. As districts tighten procurement and educators redesign assessment, the dividing line is becoming guided practice, visible reasoning, and unaided transfer—not access alone.

AI in education is separating into two paths: tools embedded in teacher-led learning sequences and general assistants left to learners to use on demand. The clearest contrast this week is between Google's reported Guided Learning trial, where educators remained central, and Tennessee evidence that optional Khanmigo use was thin. [1, 2]

Embedded tutoring is showing more promise than optional access

Google DeepMind and Fab AI report an eight-week randomized trial involving 1,763 junior-secondary students across 12 schools in Sierra Leone's Port Loko District. Guided Learning students gained 0.258 standard deviations in mathematics against a control group—translated by the report into roughly 1.2–1.7 years of typical learning progress achieved during the eight-week trial. [1]

The interaction data fit the product's intended pedagogy: across more than 113,000 exchanges, the company says students were pursuing conceptual understanding in 91.4% of conversations; Gemini asked scaffolding questions in 76% of messages and gave direct solutions in 2%. Those are process indicators, not proof of durable learning. The result also needs qualification: students who entered with stronger mathematics skills benefited most, and Google says further preregistered trials are underway in other countries. The post discloses support from Google.org and the Gates Foundation. [1]

The Tennessee evidence supplies the counterweight. In a randomized experiment across 18 middle schools, students using Khan Academy made faster mathematics gains by year two, but they used Khanmigo on only about one-third of Khan Academy days; many interactions were off-topic or attempts to obtain answers. In a separate experiment, an AI chatbot appeared without being requested and required a correct answer before students could progress. Those students moved more slowly, answered more accurately, and were slightly more likely to solve similar problems a week later without AI—a preliminary result, not a general license for chatbot use. [2]

The implication for product design is narrow but important: access is not the intervention. Prompt timing, required effort, teacher integration, and an unaided transfer check are part of the intervention. Google says its Classroom integration, planned for this fall where administrators enable Gemini, will let teachers assign course-grounded Guided Learning conversations and review class- and student-level insights, misconceptions, and secure transcripts. [3]

Sal Khan's own postmortem makes the adoption problem plain: only about 10–15% of students were willing to experiment deeply with Khanmigo, while many others needed more proactive support. The planned response is to require students to explain their reasoning, not merely select an answer, and to integrate the tutor more deeply into lessons. [4]



Sal Khan, Khan Academy: Why He Chose to Disrupt Himself With AI (16:51)

Distribution is scaling through different governance models

Google is putting a study environment directly in front of eligible students. Its Gemini student hub offers study notebooks, flashcards, and practice quizzes; notebooks can use uploaded class materials to create a study plan, diagnose knowledge gaps, generate short lessons and quizzes, and show progress on a dashboard. The offer is free for one year through Dec. 31, 2026. These are useful workflow capabilities, but the posts describe a product rollout rather than a learning outcome. [5, 6, 7]

OpenAI is taking a more district-mediated route. It says ChatGPT for Teachers is expanding to 55 additional school systems across 20 states, bringing the program to more than 100 K–12 organizations in 30 states and more than 300,000 educators and staff. Access remains limited to verified U.S. K–12 educators and is free through June 2028. OpenAI also describes a 16-state privacy framework, managed workspaces with role-based controls, and hands-on implementation support. [8]

The distinction matters. A direct-to-student hub maximizes reach and convenience; a managed teacher workspace adds institutional control and training. Neither substitutes for evidence that students retain and transfer what the tool helped them produce—the problem exposed by the Khanmigo study. [2]

Procurement is becoming an evidence-and-transition problem

The U.S. Department of Education’s original guidance rejects a simple technology-versus-no-technology debate. It asks states and districts to distinguish recreational from instructional use and to make technology purposeful, instructionally sound, and tied to measurable outcomes. Its operational advice is more demanding: evaluate demonstrated learning outcomes rather than screen time, use evidence—including randomized trials—in procurement and renewal, review implementation continuously, and support professional learning. [9]

New York City shows what happens when governance moves faster than operational planning. After backlash to its initial AI policy, the Education Department froze software purchases while it developed new guidance. With the school year approaching, principals reported being unable to renew even non-AI tools such as electronic gradebooks, family-communication platforms, digital hall passes, and reading and math software; one former department technology expert warned that schools would have to scramble in ways that undermine quality pedagogy. [10]

The city is also canceling the T-Mobile internet component of a four-year student-connectivity contract. The service cost roughly \$3.5 million per month and was projected at about \$200 million over four years; officials say existing resources can cover students who need access more cheaply. Schools were told

to identify those students and move to a Verizon program costing \$10 per student per month. Usage was uneven: one Manhattan principal said about one in five students used the devices at home, while a Brooklyn school had dozens unused and had reduced Chromebook use in class because technology was getting in the way of learning. [11]

The lesson for buyers is not to stop buying technology. It is to pair outcomes-based scrutiny with a transition plan: a district needs a way to test, renew, replace, or remove a tool without disabling the basic systems teachers and students rely on.

Assessment is moving from the submitted artifact to the learner's thinking

A New York City English and special-education teacher describes a widening gap between polished take-home writing and students' handwritten, in-class ability to organize ideas, support claims, and sustain analysis. The concern is not only cheating; it is that the assignment's cognitive practice has been outsourced while the finished product still looks successful. [12]

The teacher's redesign is sequential rather than simply anti-AI. Students annotate independently, discuss ideas, gather evidence, and form an original claim before opening a computer. They then use AI to generate counterarguments, identify gaps, suggest evidence, or offer alternative interpretations, comparing those outputs with their own and making the final decision themselves. Handwritten planning, printed-text annotation, technology-free discussion, and occasional device-free writing are used to rebuild cognitive endurance. The teacher reports deeper questions, stronger defenses of reasoning, and more thoughtful revision. [12]

A parallel response is emerging around academic-integrity enforcement. Christian Miller says even a detector reporting 99.99% reliability would not, by itself, meet the evidentiary bar for punishment or preserve classroom trust. He favors a low-confrontation conversation in which students explain their process, followed where appropriate by a reduced grade, revision, and a second chance. He also describes oral defense of code as a practical check, while acknowledging the tension with students who need technology for dysgraphia, dyslexia, or written expression. [13]

Higher education and workforce learning are shifting toward judgment

Andrew Ng argues that university curriculum processes move too slowly for AI-driven changes in work, leaving some programs preparing students for outdated roles. He advises students to supplement degrees with current online learning, while also warning that AI can raise homework scores and reduce long-term retention when it performs the work. His new Learn Vector or-

ganization is aimed at more individualized, one-to-one learning experiences. [14]



Andrew Ng: The Biggest Opportunities in AI Aren't Where You Think (14:04)

The skill response is not “learn less because the agent can code.” Ng’s AI Engineering Skills Map says developers still need fundamentals to steer agents and recognize tradeoffs in latency, reliability, security, maintainability, and cost. A higher-education workforce analysis makes the parallel institutional case: AI may make learning more adaptive, but programs must structure the experience, evaluate capability, work directly with employers, and teach judgment, communication, ethical reasoning, and general AI literacy alongside technical expertise. [15, 16]

This is also a continuing-education problem. More than half of postsecondary students reportedly enroll in at least some distance education, and one current model calls for disciplinary units to retain academic ownership while shared infrastructure supports online, adult, workforce, and lifelong learners through flexible scheduling, stackable credentials, employer partnerships, and advising. [17]

What This Means

- **For K–12 leaders:** Treat an AI tutor as a teacher-led intervention, not a software license. Pilot it with defined learning goals, required student effort, unaided transfer checks, and subgroup analysis; the current evidence shows both thin voluntary engagement and a meaningful equity qualifica-

tion. [2, 1]

- **For assessment leaders:** Design assignments so the student’s reasoning exists before AI enters the workflow, then use AI for critique or extension rather than invisible completion. Preserve documented accessibility accommodations and use process evidence, explanation, or oral defense instead of treating handwriting or a detector as a universal solution. [12, 13]
- **For edtech buyers and vendors:** Put the Department’s questions into contracts: the problem being solved, intended users, evidence of impact, implementation support, data protections, review points, and a credible exit path. New York City’s freeze shows that a policy pause without a transition plan can disrupt non-AI services as well as AI tools. [9, 10]
- **For higher-ed and L&D teams:** Pair AI fluency with domain fundamentals, judgment, and communication; build shorter, stackable learning that can be refreshed as work changes, while keeping employers involved in defining competencies. [15, 16, 17]

Watch This Space

- **Replication and distributional effects:** Google says additional preregistered trials are underway, but its first report says stronger mathematics students benefited most. The next evidence should show whether guided tutoring travels across contexts and closes—or widens—gaps. [1]
- **Outcome-based procurement in practice:** Watch whether districts turn the Department’s evidence language into renewal and removal decisions, and whether New York City can publish clear AI guidance without repeating the disruption caused by its software freeze. [9, 10]
- **Student AI at scale:** Gemini’s study hub and district-managed ChatGPT workspaces are expanding different access models. The meaningful measure will be active engagement and unaided performance, not the number of accounts provisioned. [6, 8, 2]

Sources

1. Gemini’s guided learning: results from a randomized controlled trial in Sierra Leone
2. The problem with an AI tutor: Students don’t want to use it
3. Google Classroom’s Approach to Educator-Led AI
4. Sal Khan, Khan Academy: Why He Chose to Disrupt Himself With AI
5. X post by @GeminiApp
6. X post by @GeminiApp
7. X post by @GeminiApp
8. Bringing ChatGPT for Teachers to more U.S. school districts | OpenAI
9. UNITED STATES DEPARTMENT OF EDUCATION OFFICE OF ELEMENTARY AND SECONDARY EDUCATION

10. NYC schools head toward first day without gradebooks and other software amid tech freeze
11. Eric Adams inked a \$200 million internet deal for student Chromebooks. Mamdani is canceling it.
12. Protecting Student Cognition in the Age of AI
13. AI and Academic Honesty: The Second Crisis in Schools
14. Andrew Ng: The Biggest Opportunities in AI Aren't Where You Think
15. X article by @AndrewYNg
16. Higher Education's Next Workforce Imperative
17. Continuing Education Should Be the University's Operating System, not a Silo