

AI Tutors Run Into a Dosage Problem as Evidence Starts Showing Up in Contracts

AI in EdTech Weekly

2026-10-26

AI Tutors Run Into a Dosage Problem as Evidence Starts Showing Up in Contracts

By AI in EdTech Weekly • October 26, 2026

New trials show that AI tutors only help when students actually use them, and that designs which hold back answers matter. Meanwhile, an audit finds every AI-in-education meta-analysis it examined to be flawed, and buyers in New York City, state agencies, and individual districts are tightening the rules on what gets into classrooms.

Last week's briefs asked whether AI tools can show they work and whether students use them to think. This week's evidence makes a narrower point: the tool's design matters, but whether students use it at all matters more. The organizations buying these tools are starting to write that lesson into policy and contracts.

AI tutors: good design helps, but students have to use them

A tutoring research panel hosted by Justin Reich offered the clearest data so far this fall.

- **Khanmigo in Tennessee.** A randomized study added Khan Academy's AI tutor to remedial math practice. The gains, around 0.15 standard deviations on state tests by the second year, were roughly the same as Khan Academy produced before AI. The researcher's explanation was that students rarely used the tutor: the median student opened it after only about 17% of mistakes [1]. In a follow-up one-hour trial with 7,000 students, every mistake triggered a required AI walkthrough. Students moved more slowly but were more likely to get the next question right, and slightly more likely a week later. The researcher called this a proof of concept, and concluded that evidence AI adds value on top of these platforms is "not yet" there [1].

- **An AI tutor that starts from a diagnosed misconception.** The ED platform runs a diagnostic question first, then opens a chat aimed at the specific error the student made [1]. In a 165-student trial with a human reviewing every AI message, 76% of AI drafts went out unchanged and no harmful messages reached students [1]. Students got the next question right 66% of the time with the supervised AI, 60% with a human tutor, and 56% with a static hint [1]. A trial ten times larger, with about 1,200 students, reports on October 29. A full-year UK trial with 15,000 students is now running [1].



Separating Evidence From Hype: The Latest Research in High-Impact Tutoring (40:54)

- **Why answer-giving hurts.** Panelists again pointed to the Wharton study in Turkey. Open ChatGPT access raised practice scores 48% but lowered unaided exam scores 17%, while a version that gave hints instead of answers avoided that harm [1]. There are limits on the other side too. Breaking a problem into too many tiny steps is another way of offloading the thinking [1], and tutors that withhold answers can frustrate students enough that they stop engaging [1].
- **Most students barely used it.** Stanford researchers ran two randomized trials with about 350 students in grades 1–5 on an AI literacy platform that called for 30 minutes twice a week. Just over half of the students working alone ever logged in, and they averaged 2–5 minutes a week [1]. The students who did log in tended to be higher-achieving and less likely

to have a disability [1]. Adding a human tutor for accountability raised the number of stories read by 70–80%, but that came to less than two hours of extra practice per student, with no detectable reading gains [1]. The researchers’ advice to schools: plan AI tutor rollouts as seriously as human tutoring, with protected time, a caring adult, and weekly checks on usage [1].



Separating Evidence From Hype: The Latest Research in High-Impact Tutoring (51:29)

For buyers: A well-designed tutor is necessary but not enough on its own. The deciding factors are things schools control: scheduled time, adult support, and usage monitoring.

The research base is weaker than the headlines suggest

Patrick O’Neil presented the audit Reich previewed last week. He judged all 14 AI-in-education meta-analyses he examined “severely problematic.” Both “AI” and “learning” were loosely defined. One meta-analysis pooled elementary PE interventions with medical-school simulations, and three counted mental-health chatbots as evidence of learning [2]. Of 46 primary studies he sampled from them, 28 were flawed or mismatched, some citing references that don’t exist. Twelve of the 14 meta-analyses failed to account for dependent effects, and none adequately checked for publication bias [2]. He also found that 65.8% of calls for papers in the field explicitly asked for research showing AI’s benefits

[2].

O’Neil also ran his own meta-analysis of 44 studies, applying strict criteria. The average effect was effectively zero, but the range of plausible effects ran from strongly negative ($g = -0.92$) to strongly positive ($+1.03$). He says that range, not the average, is the real finding [2]. Susanna Loeb added that of about 880 papers in her repository, only 20 met the bar for credible causal analysis in the first review. Almost all were in higher education, and none studied U.S. students using AI [2]. O’Neil’s practical advice is to read individual studies and treat promising approaches as experiments in your own setting [2].

Buyers start writing outcomes into contracts

Edtech Insiders’ takeaways from EDTECH WEEK in New York show where this is going. Twenty-four states now vet early-reading curricula against approved lists. Arizona, Florida, Missouri, and others are building ESSA evidence tiers into procurement and funding rules, which makes evidence a condition of entering the market [3]. South Bend, Indiana, signed a middle-school literacy contract with Really Great Reading that sets requirements for dosage, attendance, and staffing, and ties 40% of the contract’s value to students reaching proficiency on the 2027 state test [3]. Speakers named the same “5% problem” the tutoring researchers found: most students never get near the dosage a study assumed. One proposed fix is usage targets that step up over time, owned jointly by the district and the vendor [3]. For now, vendors bear most of the cost of producing evidence, which favors larger companies [3].

NYC’s AI restrictions hit more than 38 products, and some vendors don’t know why

Chalkbeat obtained New York City’s list of more than 38 flagged products. It includes widely used reading and math curricula, assessment tools, and a chatbot the Education Department built itself [4]. The rule targets student-facing generative AI features, not necessarily whole products, but the department hasn’t explained how it decided which tools qualify, including in cases where vendors dispute that their tools use AI [4]. Amira’s AI reading tutor, used in more than 200 schools, was banned outright [4]. Other vendors are adjusting feature by feature. All In Learning turned off AI-assisted grading, and Magma Math disabled AI hints on September 4 and was cleared to operate again on October 5 [4].

In Washington, policy is pulling in opposite directions. NTIA head Arielle Roth reportedly compared ed-tech to tobacco in a private meeting; she says she doesn’t remember the specific analogies [5]. She has also helped prompt a review of the \$3 billion E-Rate fund [5]. Meanwhile, the Education Department has issued relatively favorable guidance on ed-tech [5].

Teachers push back on mandated AI essay grading

A co-teacher reports that an administrator wants to require AI scoring for ELA essays. The teacher says the AI gave 80s to 100s to essays missing major components, including essays apparently written by AI [6]. In a meeting, staff regraded five AI-scored papers and found the AI had inflated grades and written feedback that didn't match what students wrote, such as critiquing a thesis that didn't exist. The administration still wants a department-wide approach [7]. Another teacher says their department dropped its mandate after the tool rated strong writing 1–2 out of 4 and gave copied paragraphs a 4 [8]. These are individual accounts, but they show that AI grading can be wrong in both directions, too generous and too harsh, which makes it hard to plan reteaching around its scores.

Teen safeguards and student self-reports

Common Sense Media ran 4,000 prompts against ChatGPT for teens and rarely triggered its protection mode. Study mode was easy to exit: students could ask for the answer or delete “@study” from their prompt [9]. Common Sense's August survey also found that 70% of U.S. teens use AI for schoolwork. Among teens who use AI, 38% said it leads them to come up with fewer ideas of their own, and 39% said they miss out on learning when they use it to finish assignments [10].

Higher ed keeps scaling access

Jisc signed a framework agreement with Google that makes Google AI Pro and Workspace for Education easier to procure for UK and Irish further- and higher-education institutions. It comes with on-campus training and free certifications [11]. As the evidence above shows, giving students access does not guarantee they will use these tools in ways that help them learn.

Worth watching

- **The ED platform's 1,200-student trial report on October 29**, and later results from its full-year UK trial with 15,000 students.
- **Whether NYC publishes its criteria** for deciding which products count as student-facing AI.
- **Whether more districts copy South Bend's outcomes-based contract**, especially for AI tools whose results depend on how much students use them.

Sources

1. Separating Evidence From Hype: The Latest Research in High-Impact

Tutoring

2. Presumed Effective: The Manufacturing of an Evidence Base for AI-in-Ed Through Flawed Meta-Analysis
3. Takeaways from EDTECH WEEK: Evidence Gets You In. Outcomes Get You Renewed. AI Gets Vetted.
4. NYC's AI ban confusion: what's allowed and what's banned in city schools
5. The Trump official who compared ed-tech to the tobacco industry
6. r/Teachers post by u/EquipmentThin9682
7. r/Teachers comment by u/EquipmentThin9682
8. r/Teachers comment by u/sincerely0urs
9. AI in Education Podcast: When AI Helps Students Learn Less: What the Research Reveals
10. How AI Is Breaking the Way We Reward High-Achieving Students
11. Jisc & Google for Education Partnership to bring AI to UK Universities