

# Guided AI Helps Learning, Answer-Giving AI Undercuts It, and Students Keep Picking the Answers

AI in EdTech Weekly

2026-10-12

## Guided AI Helps Learning, Answer-Giving AI Undercuts It, and Students Keep Picking the Answers

*By AI in EdTech Weekly • October 12, 2026*

New trials and pilots show the same pattern: AI improves learning when students use it as a tutor and hurts it when students let it do the work. Meanwhile OpenAI shipped GPT-6, removed a sharing tool many teachers relied on, and New York moved further toward screen limits.

Last week's question was whether AI tools can show they work. This week brought a partial answer: AI seems to help when students use it to think, and to hurt when they use it to skip the thinking. The hard part is that when students get to choose, many go for the shortcut.

### **The evidence: helpful as a tutor, harmful as a ghostwriter**

Ethan Mollick highlighted new randomized trials run with the older GPT-4o. AI access raised test scores, and a smaller gain was still there a week later. That gain held for students who used AI as a tutor (“augmentation”) and faded for students who had AI write for them (“automation”) [1].

Andrew Ng described the other side. He said rigorous studies are starting to show that chatbot help on homework raises homework scores while lowering final-exam scores and other measures of long-term learning. He argues that most people use chatbots in ways that are “very harmful for learning,” whether in K–12, higher ed, or adult learning [2].



*World Development Report 2026: Fireside Chat with Andrew Ng - Making AI Work for Development (24:24)*

Ng also mentioned an unpublished Stanford experiment. Students were randomly assigned to replace graded homework with a 30-minute presentation to a TA, and they did “dramatically better.” He asked right away whether large state universities, let alone schools in the developing world, could afford that model [2].

Google gave the most positive data point, though it comes from a vendor. In Sierra Leone, math students who used Gemini’s Guided Learning under their teacher’s direction gained the equivalent of 1.2–1.7 years of typical instruction on standardized tests within eight weeks. More than 90% of their interactions involved working through concepts step by step, not asking for answers [3]. Google also says teachers using Gemini for Education reported 70% less time on administrative tasks, a self-reported figure [3].

Justin Reich urged caution about the research base itself. A colleague of his reviewed 14 education meta-analyses that all found strong positive links between generative AI and learning, and judged every one “severely flawed” [4]. About 40 states have published AI guidance, but in Reich’s view none offers an evidence-based playbook. His advice is to treat anything sold as “best practice” as a hypothesis and test it locally against real student work [4].

**In practice:** Whether a tool is designed to guide students matters less than whether students actually use it that way. Teacher direction looks like the

deciding factor in the most positive results.

## **When students can choose, the guided tool loses**

Estonia’s national OpenAI program shows the uptake problem clearly. Every teacher got paid AI tools. 95.9% reported using AI professionally, and users saved just under two hours a week [5]. Students were another story. Of 200,000 invited, under two-thirds activated accounts, and only 20% of those used ITI, the Socratic tutoring app, against a 60% target [5]. Students found it slow and tedious next to consumer ChatGPT, which just gives the answer. ITI scored 3/5 on satisfaction with a net promoter score of  $-27$  [5].

Sal Khan described the same pattern with Khanmigo. In most classrooms only about 15% of students engaged deeply with a tutor that waited for them to ask questions [6]. The version launched for this school year is proactive. It asks students to reflect on both right and wrong answers so teachers get better evidence of what students actually understand [6]. Khan says the tutoring-session error rate is “well under 5%,” meaning fewer than one in 20 conversations has at least one misstep [6]. He presents Khanmigo as an added layer on Khan Academy, whose 50+ efficacy studies he says show 30–60% acceleration at 60 minutes a week. Those results come from the platform, not from the AI [6]. In West Hempstead, New York, the superintendent describes the payoff as “return on learning”: seeing students’ reasoning, where they got stuck, and how they revised, not just whether the final answer was right [7].

## **OpenAI: new interface, teen accounts, and a tool teachers are losing**

OpenAI rolled out GPT-6 with “Intelligent UI.” Responses can now mix text, charts, and interactive elements such as buttons and forms, depending on the question [8]. It went to paid tiers first, with Free and Go users getting access the next day. Paid tiers run GPT-6 Sol and free tiers run GPT-6 Luna [9]. Mollick, who had early access, called it a welcome change from “walls of text” [10]. The announcement includes no evidence about learning outcomes. Visual, interactive answers may help with exploring a topic, but they are still answers.

OpenAI also shared progress on ChatGPT for Teens. Protections turn on by default for accounts identified as belonging to under-18s. A coming College Planner will pull together application requirements, deadlines, and financial-aid steps for U.S. students aiming at four-year colleges [11]. OpenAI says it will evaluate how the protections work in practice [11].

The setback for teachers: Custom GPTs and Gemini Gems can no longer be created, and OpenAI’s existing GPTs stop working in December. Teachers used them to build shareable tools, such as a bot that gives students feedback on their writing. The suggested replacement, ChatGPT plugins, can’t be shared [5]. Mollick made the same point: he can move his GPT to a private plugin,

but he built it so other people could use it [12]. Schools that built workflows around shared GPTs should plan replacements before December.

In Australia, UNSW will give ChatGPT to about 80,000 students from next academic year. Flinders is giving Gemini to 30,000 students along with Google Career Certificates [5].

### **New York pulls back from screens, with exceptions**

New York City made its first exemption to its ban on individual devices for pre-K-2. The CUNY Reading Fellows program can keep offering virtual one-on-one tutoring. About 30% of its 1,600 students last year were tutored remotely [13]. Officials said live instruction from human teachers can qualify for exemptions, while software-delivered instruction cannot [13]. A randomized study of the program found no statistically significant difference between virtual and in-person outcomes, though virtual students attended fewer sessions [13].

Separately, more than half of City Council members asked the state to allow paper exams for grades 3-8, only months after the state fully moved to digital testing. The state says it is reviewing the request [14]. The state argues digital tests score faster and can adapt to each student's level [14], and there is little evidence that the switch caused last year's drop in reading scores [14].

### **Classrooms move to paper, at a cost**

Teacher forums show the problem directly. A high-school marine science teacher found that 15 of 22 essays were AI-written, even though the students had been warned and had other project formats to choose from. Hidden off-topic words planted in the assignment went unnoticed [15]. Many teachers are going back to paper. One teacher with three years of paper-only work reports "a large improvement in classwork" [16]. The trade-offs are real, though. One teacher's biggest complaint is the extra hand-grading [17]. Another says students still use AI to fill in paper work [18]. An AP teacher who moved all assessments into class saw end-of-quarter drops spike. That teacher regrets losing projects that once let students who test poorly show what they had mastered [19]. These are individual accounts, not survey data.

Detection tools don't solve this. OpenAI will watermark text in the EU, but the watermark only suggests whether text was "likely generated by an OpenAI model." It does not identify who wrote it [20]. Short passages often go undetected, rewriting or translating can remove the watermark, and only approved researchers can use the detector for now [21]. Google made SynthID Detector available to everyone for checking AI-generated content, including content from partners such as OpenAI and NVIDIA [22]. That makes it more useful for media literacy than for checking authorship.

## Workforce and new institutions

Launchpad Philly trains young people who mostly graduated from public schools without college degrees. Of its first 12 intensive-program graduates, who had little AI training, three work full-time a year later, two are in college, and the rest piece together freelance and part-time work [23]. The program now offers generative-AI and project-management certifications. It teaches students to explain technical tools to people in business roles, and it runs a paid project arm that pays participants \$20 an hour [23].

Khan Academy, TED, and ETS are building the Khan TED Institute. The goal is a full degree under \$10,000, built on team projects and assessed through peer and AI review of a portfolio, not GPA [6]. Khan says the interest list is approaching 30,000, mostly people who already have degrees or are in college [6].

Alpha School co-founder MacKenzie Price said her schools are no longer in the top 1% on MAP after admitting hundreds of students who averaged 2.5 grade levels behind. She claims some students grew six times faster than the national rate [24]. She expects the Timeback platform to be downloadable for other schools within the next year [24].

## Worth watching

- **The New York State Education Department’s decision** on paper testing, promised “in the coming weeks.”
- **Khanmigo’s proactive redesign:** whether it gets more than the 15% of students who engaged with the passive version.
- **What replaces Custom GPTs** for shareable classroom tools before December.
- **Reich’s upcoming webinar** on the flawed meta-analyses, which could change how districts read vendor evidence.

---

## Sources

1. X post by @emollick
2. World Development Report 2026: Fireside Chat with Andrew Ng - Making AI Work for Development
3. Building Responsibly: The Learning Science Behind Google’s Learning Tools
4. The Homework Machine: Navigating Uncertainty and the Arrival of AI | Justin Reich
5. The AI Learning Paradox: Better Results, Less Learning?
6. SAL KHAN on AI Tutors, Durable Skills & the Khan TED Institute | Trending in Ed w Mike Palmer
7. From ROI to “Return on Learning”: A District’s Khan Academy Story

8. X post by @OpenAI
9. X post by @OpenAI
10. X post by @emollick
11. X post by @OpenAI
12. X post by @emollick
13. Mamdani's school device ban has its first exception: virtual tutoring
14. After yearslong transition to digital tests, state officials weigh return to paper exams for NYC
15. r/Teachers post by u/Superior\_Sass
16. r/Teachers comment by u/Comprehensive\_Tie431
17. r/Teachers comment by u/MycologistPutrid7494
18. r/Teachers comment by u/ColorfulEgg
19. r/Teachers comment by u/Potential\_Fishing942
20. X post by @OpenAI
21. X post by @OpenAI
22. X post by @GoogleDeepMind
23. AI forces coding program to revamp how it prepares Philadelphia students for tech jobs
24. School Series, ep1 | Alpha School, conversation with Cofounder MacKenzie Price