BROOKINGS

COMMENTARY

What the research shows about generative AI in tutoring

Mary Burns

January 27, 2026

- → While tutoring platforms enhanced by generative AI introduce new concerns around accuracy, pedagogical judgment, and possible dependence, the evidence shows that these platforms can hold numerous benefits for students—if designed responsibly.
- → It's not just students who benefit. Tutoring that incorporates generative AI promises substantial benefits for teachers and education systems as a whole.
- → But in order to work as intended, safeguards must be built into the AI platforms. Areas for improvement remain, including enhancing personalization, fine-tuning effective feedback, and prioritizing a hybrid approach where human teachers and AI work together.

Of AI's many educational possibilities, perhaps none holds more promise and excitement than its potential to teach—or more precisely, to tutor—individual students.

Early iterations of tutoring, in the form of Intelligent Tutoring Systems, have long captured the imagination of policymakers, educational planners, and researchers. High-quality research has demonstrated that Intelligent Tutoring Systems (ITSs) can match the success of human tutoring. Nesbit et al. (2014 (https://www.researchgate.net/publication/286583493_How_Effective_are_Intelligent_Tutoring_Systems_in_Computer_Science_Education)) found "a significant advantage of ITS over teacher-led classroom instruction and non-ITS computer-based instruction." Kulik and Fletcher's (2016 (https://doi.org/10.3102/0034654315581420)) meta-review of 50 studies concludes that ITSs can "match the success" of human tutoring while Major et al. (2021 (https://bera-journals.onlinelibrary.wiley.com/doi/epdf/10.1111/bjet.13116)) report that tutoring

platforms can support tailored, developmentally appropriate "teaching at the right level" approaches.

The integration of generative AI in tutoring systems has amplified this enthusiasm for technology-based tutoring. Traditionally, intelligent tutoring systems have relied primarily on symbolic (https://www.thehackettgroup.com/glossary/symbolic-ai/) and rule-based (https://www.geeksforgeeks.org/artificial-intelligence/rule-based-system-in-ai/) AI approaches, often combined with machine learning (https://www.ibm.com/think/topics/machine-learning) components for student modeling and instructional decisionmaking. Generative AI is now beginning to address longstanding limitations in traditional intelligent tutoring systems and supplant some of its traditional components (particularly rule-based dialogue, feedback, and content generation) while reshaping how tutoring systems are designed and deployed. In this blog, we examine the educational implications of this shift, highlighting the impact of tutoring platforms enhanced by generative AI and how this integration is adding value to such platforms.

Generative AI as tutor: The evidence for effectiveness

Thus far, many claims about the educational benefits of generative AI have outpaced high-quality evidence. But a number of recent rigorous studies suggest that tutoring systems that integrate generative AI can perform many of the core functions traditionally handled by human beings or expert-authored scripts, deliver learning gains and efficiency, and offer a path to solve the challenge of individualized education at scale.

Table 1 summarizes four recent randomized controlled trials outlining the role of generative AI in tutoring platforms.

TABLE 1

The role of generative AI in tutoring

Study	Summary	Key aspects of successful tutoring	Evidence for effectiveness
U.S.: Harvard University Physics classes (Kestin et al. 2025)	• **Randomized Controlled Trial (RCT)** involving 194 undergraduate physics students at Harvard University. • Compared an AI tutor against active learning classroom instruction, with both approaches designed using identical pedagogical best practices.	• **Adaptive pacing/self-pacing:** Students can take the time needed to build conceptual understanding or move ahead if familiar with the material. • **Pedagogical fine-tuning:** The AI was developed using content-rich prompt engineering based on pedagogy and educational psychology. • **Accuracy safeguards:** Prompts were enriched with detailed step-by-step answers to guide the AI and ensure high-quality, accurate explanations.	• Students using the AI tutor (treatment group) achieved **more than double the learning gains** relative to the pre-test baseline compared to the active lecture group. • The treatment group achieved a higher median post-test score (4.5 vs. 3.5) while spending **less time on task** (median 49 minutes vs. 60 minutes). • The treatment group reported feeling significantly **more engaged and motivated.**
U.K.: (LearnLM Team, Google & Eedi, 2025)	• **An exploratory RCT** with 165 students across five UK secondary schools focused on mathematics. • The study rigorously compared static hints, human tutoring, and LearnLM (a pedagogically fine-tuned generative AI model) and was directly supervised by expert human tutors.	• **Socratic dialogue:** LearnLM was prompted to adopt a Socratic approach to guide students to identify their own mistakes. • **Prompted inquisitive, student-led interaction**, according to tutors. • **Hybrid vigor/human supervision:** Expert tutors reviewed and retained full control to edit or replace LearnLM's drafts, ensuring social-emotional nuance and appropriate pacing. • **Pedagogical fine-tuning** by programmers and tutors plays a role in delivering effective, individualized tutoring at scale.	• Students guided by the supervised LearnLM were **5.5 percentage points more likely to solve novel problems** on subsequent topics than those who received tutoring from human tutors alone, demonstrating **better knowledge transfer.** • Supervised LearnLM proved **just as effective as human tutors** at helping students correct mistakes and resolve misconceptions (95.4% success rate vs. 94.9% for human tutors). • **Students using the supervised AI tutor performed slightly better** than those who texted human tutors and were able to solve new kinds of problems on subsequent topics successfully 66.2% of the time, compared to 60.7% with human tutors.
		• **Adaptive personalization:** The study measured and compared different approaches to	• The intervention demonstrated a **significant improvement of 0.31**

	• RCT deployed	personalization, examining how adaptive approaches dynamically adjust instructional content, pacing, or difficulty based on student	**standard deviation** on an assessment that included English topics aligned with the Nigerian curriculum,

Tutoring platforms enhanced by generative AI introduce new concerns around accuracy, pedagogical judgment, and possible dependence (Bastani et al., 2025 (https://www.pnas.org/doi/10.1073/pnas.2422633122)). But as the empirical evidence from Table 1 shows, these platforms hold numerous benefits for students. These include substantial learning gains across all studies, greater knowledge transfer, improved motivation and engagement, and efficiency. Students can be saved from classes "pitched to the median" in which high-achieving learners are insufficiently challenged while struggling students fall behind (The Economist, 2025, 11 (https://www.economist.com/weeklyedition/2025-12-06)). Critically, this enables the kind of "private tutors, personalized syllabus and bespoke learning" opportunities that were previously "available only to the privileged few."

> " As the empirical evidence from Table 1 shows, these platforms hold numerous benefits for students."

A new standard: Where AI can add value to existing tutoring systems

As noted, tutoring systems come with a long body of evidence for their effectiveness, but the integration of generative AI adds distinctive value to these platforms in several important ways, which as seen in Table 1 contributes to the success of these profiled tutoring programs. First, large language models can generate naturalistic dialogue such as explanations tailored to individual student questions, rather than selecting from pre-scripted responses typical of rule-based tutoring systems. Students can ask follow-up questions in natural language and receive contextually appropriate answers, and tutoring platforms powered by generative AI can provide sophisticated feedback on open-ended responses, particularly in domains like writing or mathematical problem-solving. This addresses one of the historical limitations of rule-based tutoring systems, which struggle with student queries that fall outside their predetermined pathways.

Second, while earlier ITS platforms could implement predetermined question sequences or select from question banks based on student performance, generative AI enables tutoring systems to formulate questions dynamically based on what

students actually write and say. This can involve the use of a greater variety of questioning techniques—probing questions, clarifying questions, Socratic questions, a mix of open and closed questions—and the deployment of certain support techniques like providing hints, scaffolding, and guidance.

Third, tutoring platforms enhanced by generative AI help to create psychologically safe learning environments where students feel comfortable seeking the academic help they need. For example, the naturalistic language capabilities of these platforms enable students to express confusion, ask clarifying questions, and articulate partial understanding in their own words without fear of judgment. Unlike classroom settings where students may be embarrassed or hesitate to reveal gaps in their knowledge in front of peers, or even one-on-one human tutoring, AI tutors respond with infinite patience and non-judgmental support.

Finally, tutoring platforms are versatile. They can be integrative, used during class along with teacher instruction, or substitutive, used as a substitute for teacher instruction. Generative AI amplifies that versatility. These platforms can be accessed in school, at home, or after school, with teacher support, alone or with classmates, as self-paced instruction or small-group collaboration, and students can personalize the type of support they need (flashcards, practice sentences, games) depending on the design of the platform. The individualized nature of AI tutoring confers privacy from peers while the asynchronous availability of these systems allows students to engage with material at the time, pace, and place of their choice—seeking help when they feel ready to learn. And assuming accuracy rates of LLMs improve, students don't need access to expensive tutoring platforms to learn.

Whole-system benefits

But it's not just students who benefit from these tutoring platforms. If designed well, tutoring programs that incorporate generative AI promise substantial benefits for teachers and education systems as a whole.

For teachers, such tutoring programs could significantly increase productivity (Keppler et al., 2024 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4924786)), particularly in under-resourced schools where classes are large, time is limited, and human resources are scarce—precisely the contexts where students, and their teachers, need the greatest support. These platforms are especially valuable for novice teachers, who are often assigned to challenging environments despite lacking

pedagogical expertise in lesson design, assessment construction, and differentiated instruction. Tutoring systems enhanced by generative AI can provide scaffolding, templates, and real-time guidance that help novice teachers deliver more effective instruction while building their professional skills. Like Interactive Audio Instruction (https://edc.org/sites/default/files/edc-distance-edu-book-CH2.pdf) (IAI) and TV-based classroom instruction have long done, these AI-enhanced platforms can serve as a form of "dual audience direct instruction (https://documents1.worldbank.org/curated/en/762911468281358815/pdf/466210NWP0Box31Radio0Instruction111.pdf#page=7) ," educating unqualified or volunteer teachers as they educate students.

Beyond supporting teacher development, these platforms can restructure how instructional time is used through models like flipped classrooms where students engage with introductory material at home via tutoring platforms enhanced by generative AI. This frees up precious in-class time for teachers to facilitate high yield tasks, such as collaborative work, critical thinking, and project-based learning. This approach allows teachers to focus their expertise on the complex work of guiding student discourse and application, rather than spending the majority of class time on content delivery—maximizing the impact of scarce instructional time while the platform carries out tasks that might otherwise consume the entire class period.

At the system level, tutoring platforms that leverage generative AI can enable dynamic content creation, producing varied and on-demand practice problems, question banks, and hint sequences. This eliminates the need for extensive pre-authoring of every possible instructional scenario and developing extensive question banks with varying difficulty levels to accurately categorize students into performance tiers—a requirement that Rodriguez-Segura (2022 (https://openknowledge.worldbank.org/server/api/core/bitstreams/6ac08b1a-d072-4727-8678-d42bbee86a8a/content)) emphasizes is essential for traditional adaptive software. This significantly reduces the personnel and development costs and time required to build tutoring systems for new domains.

The cost-effectiveness of tutoring platforms enhanced by generative AI outlined in Table 1 (particularly in both the World Bank and Stanford studies) make them promising candidates for large-scale implementation (DeSimone et al., 2025, 22 (https://openknowledge.worldbank.org/entities/publication/15e1ff08-15ae-4f7a-b2a8-d146e6c113ee#page=22)). This potential for scalability is particularly encouraging for countries grappling with severe teacher shortages, high population growth, and increasing teacher attrition rates, especially those seeking affordable, efficient ways to

address learning gaps in resource-constrained environments. The need for specialized knowledge and teacher shortages at the secondary level make "teaching at the right level" even more difficult to implement. By harnessing these tutoring platforms to provide personalized, adaptive learning at scale, governments can take decisive steps toward improving learning outcomes in contexts that have traditionally faced significant educational challenges (DeSimone et al., 2025, 23 (https://openknowledge.worldbank.org/entities/publication/15e1ff08-15ae-4f7a-b2a8-d146e6c113ee#page=23)), with the potential to scale expert-level support and deliver high-quality experiences where they are most needed (Wang et al., 13 (https://nssa.stanford.edu/studies/tutor-copilot-human-ai-approach-scaling-real-time-expertise)).

Design matters

But the integration of generative AI into tutoring platforms doesn't apply such techniques automatically—it must be designed to do so. While each tutoring platform outlined in Table 1 has its own particular affordances, all share a focus on pedagogically sound design, an approach that my new co-authored Brookings publication "A new direction for students in an AI world: Prosper, prepare, protect (https://www.brookings.edu/articles/a-new-direction-for-students-in-an-ai-world-prosper-prepare-protect/) " has shown is critical for AI to deliver its promised learning benefits.

Studies have shown that students using generative AI without specific guidance demonstrate very "limited reflection" on learning material (Krupp et al. 2023, 8 (https://arxiv.org/pdf/2309.03087#page=8)). The success of these tutoring platforms enhanced by generative AI is grounded in pedagogies that "teach, not tell"—employing Socratic approaches, asking questions that guide students to identify their own mistakes rather than providing direct answers, and breaking down complex issues to manage cognitive load and make learning more digestible. Content is highly scaffolded and organized sequentially to guide students through problems rather than presenting them all at once. In the Harvard study, explicit safeguards, such as step-by-step solutions, are embedded in prompts to ensure the generative AI component delivers high-quality explanations and prevents hallucinations. And, as will be further discussed, in three of these examples, generative AI doesn't replace human beings: rather, it complements their role.

There is still room to improve

Evidence strongly supports the effectiveness of well-designed tutoring platforms enhanced by generative AI, such as those profiled in Table 1. When designed according to sound pedagogical principles, these platforms can deliver learning gains and efficiency, offering a path to solve the challenge of providing quality, individualized education at scale.

Yet beyond these exemplary cases, there is still room to enhance the learning potential and pedagogical possibilities of AI tutoring programs generally–whether they incorporate generative AI or rely on traditional intelligent tutoring systems. Below are three areas where tutoring programs, including those incorporating generative AI, could broadly improve:

True personalization vs. individualization

While some tutoring platforms genuinely personalize student learning, many others merely individualize it but call it personalization. This distinction matters considerably. As ISTE (2018) notes, individualized learning involves students progressing through the curriculum at different speeds based on their particular learning needs. This approach serves students (https://iste.org/blog/personalized-vs-differentiated-vs-individualized-learning) who need to review previously covered material, those who want to skip content they've already mastered, and those who need more time with particular topics to develop understanding. Individualized approaches fine-tune problems (https://iste.org/blog/personalized-vs-differentiated-vs-individualized-learning) based on difficulty, promote self-paced learning, and move students through pre-structured content at individualized rates.

In contrast, personalized learning (https://iste.org/blog/personalized-vs-differentiated-vs-individualized-learning) is tailored to the preferences and interests of various learners and paced to each student's unique needs. Academic goals, curriculum, and content—as well as method and pace—can all conceivably vary in a personalized learning environment. Unlike individualized instruction, personalized learning involves the student in the creation of learning activities and relies more heavily on personal interests and innate curiosity. True personalization still remains a challenge for many AI tutoring platforms.

Feedback

Feedback is an essential element of effective tutoring programs, but its presence alone does not guarantee effectiveness. Its impact depends on several factors, (https://www.edutopia.org/article/giving-teachers-better-feedback/) including the motivation of individuals involved, the timing of feedback, and the nature of the feedback itself. Many tutoring platforms give feedback at the answer level (right/wrong), but feedback should occur at both the answer level and the step level. Effective feedback includes two essential elements: (https://www.edutopia.org/article/giving-teachers-better-feedback/) verification (assessing the correctness of a student response) and elaboration (offering information to fill gaps in understanding and guide the recipient toward a desired approach). It must be "high information"—containing specific actionable information that helps learners meet their goals—and "forward focused"—outlining what learners can do versus what they've already done.

Hybrid vigor

Finally, for many, one attraction of tutoring systems—those using generative AI and more traditional platforms—is that they replace teachers. But as the studies in Table 1 suggest, the optimal tutoring model appears to be one of human-AI hybrid vigor, where teachers continue to play "an essential role in monitoring and guiding students' use of the LLM to ensure it [is] used appropriately and productively" (DeSimone et al, 2025, 21 (https://documents1.worldbank.org/curated/en/099548105192529324/pdf/IDU-c09f40d8-9ff8-42dc-b315-591157499be7.pdf#page=21)) and where tutoring programs liberate teachers from teaching content to capitalize on classroom time for "activities and projects that foster advanced cognitive skills such as critical thinking and content synthesis." (Kestin et al. p 6 (https://www.nature.com/articles/s41598-025-97652-6)) As the above studies from Nigeria, the U.K., and the U.S. show, the educational success of tutoring platforms, those incorporating generative AI and those that do not, lies in the thoughtful collaboration of humans and AI in the learning process.

AUTHOR

Mary Burns **Consultant** - Brookings Global Task Force on AI in Education

---