

EVIDENCE • WHY THIS WORKS

The evidence behind AI-assisted coaching

Two landmark randomised controlled trials on what happens when AI is designed to deliver expert, personalised feedback in education, and what they tell us about the approach Starlight is built on.

Starlight is built on a simple idea with deep research support: teaching improves fastest when feedback is specific, timely, actionable and regular. The harder question has always been how to deliver feedback of that quality at scale. Two rigorous, independent studies, from Harvard and Stanford, now offer part of the answer. Both are randomised controlled trials, the strongest form of effectiveness evidence, and both point the same way: when AI is designed around sound pedagogy and used to put expert guidance in an educator's hands, the effects are large.

What this evidence is, and what it is not

The studies below evaluated other AI tools, not Starlight. They are evidence for the approach Starlight is built on, not measurements of Starlight itself. Starlight's own independent effectiveness research is in development, and where our evidence stands today is set out at the end of this page.

Stanford: AI that makes educators better

900 tutors	1,800 students	+4 pts topic mastery	+9 pts lower-rated tutors
----------------------	--------------------------	--------------------------------	-------------------------------------

Tutor CoPilot, developed at Stanford, puts expert-style pedagogical guidance in front of tutors while they teach. In the first randomised controlled trial of a human-AI system in live tutoring, 900 tutors and 1,800 K-12 students from historically under-served communities took part. Students working with tutors who had access to the tool were four percentage points more likely to master a topic. The effect was largest where it mattered most: for students of the lowest-rated tutors, mastery rose by nine percentage points.

Analysis of more than half a million tutoring messages showed why. Tutors using the tool were more likely to reach for high-quality moves, such as asking guiding questions, and less likely to simply hand over the answer. The researchers are candid about limitations, including AI suggestions that were sometimes not pitched at the right grade level. The headline holds, though: AI can raise the quality of a human educator's practice, at scale, with the biggest gains for those who have least access to expert support. [1]

Harvard: design is what makes AI work

194 students	2× learning gains	Less time on task	Higher engagement
------------------------	-----------------------------	-----------------------------	-----------------------------

The second study, published in Scientific Reports, tested a purpose-built AI tutor against active learning, one of the most effective classroom methods there is, in Harvard's largest physics course. Students randomly assigned to the AI tutor learned more than twice as much as those in the active-learning class, in less time, and reported higher engagement and motivation.

The authors are emphatic about the reason, and it is the part most relevant here. The gains did not come from the raw power of the model. They came from careful pedagogical design: scaffolding, one step at a time, immediate personalised feedback, and self-pacing. A generic chatbot does not produce these results; a well-designed one does. The study is small and short,

and concerns undergraduates learning physics rather than teachers developing their practice. Read for what it establishes, its message is clear: the value of AI in education lies in the pedagogy built into it. [2]

What this means for Starlight

Read together, the two trials make a consistent case. AI produces real gains in education when it is designed around evidence-based pedagogy, as at Harvard, and used to scale expert guidance to the human educator, as at Stanford. That is precisely the space Starlight occupies. Where Tutor CoPilot guides a tutor mid-session and Harvard's tutor teaches a student directly, Starlight analyses a whole recorded lesson and gives the teacher private, structured coaching grounded in what actually happened in the room, feedback designed to be specific, timely, actionable and regular.

We are careful about what this does and does not show. These are studies of tutoring, not of teacher coaching from lesson transcripts. They establish that the mechanism works: expert-quality feedback, delivered at scale, changes what educators do and improves outcomes. That mechanism is the one Starlight is built to provide for teacher development. It is the reason we designed the platform the way we did.

Where Starlight's own evidence stands

We hold ourselves to the standard we would want any school to apply. Using the widely recognised ESSA evidence framework, Starlight's own evidence today sits at the 'demonstrates a rationale' level: a clear, research-grounded logic for why it works, supported by the practice-level findings in our school case studies, where documented teaching behaviours strengthened measurably over a year. We have made a public commitment to reach the 'promising' tier through an independent correlational study, now in development. We will report what it finds, whatever it finds.

See it for your school

Explore how Starlight brings this kind of feedback to every teacher, and read what it looks like in a real school.

starlightmentor.com/demo-request

Spark Insight with Starlight, and put research-grounded coaching within every teacher's reach.

References

- [1] Wang, R. E., Ribeiro, A. T., Robinson, C. D., Loeb, S., & Demszky, D. (2024). Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise. arXiv:2410.03017. arxiv.org/abs/2410.03017
- [2] Kestin, G., Miller, K., Klaes, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports, 15, 17458. doi.org/10.1038/s41598-025-97652-6
- [3] Stanford EduNLP Lab. Tutor CoPilot project page. edunlp.stanford.edu/projects/tutor-copilot

Starlight by STAR21 · info@starlightmentor.com · www.starlightmentor.com

The independent studies cited evaluate other AI tools and do not measure Starlight. They evidence the approach Starlight is built on.