

# AI Tutoring and the Efficacy Question

*What randomized evidence says AI tutoring actually does to learning — and what the market is paying for*

Scaffold — Research Desk

Coverage: Scaffold

## ABSTRACT

Does AI tutoring actually improve learning, or is it a better interface wrapped around an unproven claim? This report weighs the randomized evidence — the Stanford Tutor CoPilot RCT ( $\approx 900$  tutors, 1,800 K-12 students), Google's LearnLM UK classroom RCT (165 students), and the Khanmigo trials — against the techno-economic reality of the products being sold, then against the four-decade ghost of Bloom's "two-sigma" claim. The honest finding: the best-evidenced AI tutoring today is human-in-the-loop, where AI raises a weak tutor toward an average one (Tutor CoPilot: +4 percentage-point mastery overall, +9 pp for the lowest-rated tutors, at  $\sim \$20$ /tutor/year), not an autonomous tutor delivering two sigma. Fully-autonomous AI tutoring matches human tutors on near-term task measures (LearnLM: 93.0% vs 91.2% second-attempt accuracy) but lacks durable, independently-audited learning-gain evidence. Implication: efficacy is real but bounded and unevenly distributed; the market is pricing the hype curve, not the evidence curve. We frame the exposure for DUOL, COUR, CHGG and PSO, with GOOGL as the platform layer. Not investment advice.

*Keywords: AI tutoring, LearnLM, Tutor CoPilot, Khanmigo, two-sigma problem, learning efficacy, edtech economics, randomized controlled trial, effect size, DUOL, COUR, CHGG, PSO*

## Executive Summary

The question this report answers is narrow and decision-grade: **does AI tutoring measurably improve learning, and if so, by how much, for whom, and at what cost?** The honest answer separates three things the market routinely conflates — a *better interface*, a *real but bounded learning effect*, and the four-decade fantasy of *two-sigma* gains.

The strongest causal evidence today is for **human-in-the-loop** systems, where AI coaches a live tutor in real time. Stanford's Tutor CoPilot randomized controlled trial (RCT) —  $\approx 900$  tutors and  $\approx 1,800$  K-12 students from under-served communities — found students of AI-assisted tutors were **4 percentage points (pp) more likely to master a topic ( $p < 0.01$ )**, rising to **+9 pp for students of the lowest-rated tutors**, at a cost of roughly **\$20 per tutor per year** [1][2]. The mechanism is observable: assisted tutors were  $\sim 10$  pp more likely to prompt students to explain their reasoning rather than give the answer [1].

**Fully-autonomous** AI tutoring is now competitive with human tutors on *near-term task* measures. Google DeepMind's LearnLM RCT in five UK secondary schools (165 students, May–June 2025) found the supervised AI matched or slightly beat human tutors on second-attempt accuracy (**93.0% vs 91.2%**) and on knowledge transfer to a new topic (**66.2% vs 60.7%**, 93.6% posterior credibility), with **zero harmful messages** and a 0.1% factual-error rate across 3,617 drafted messages [3]. But those are session-level proxies, not durable, externally-audited learning gains.

The two-sigma frame — Benjamin Bloom's 1984 claim that one-to-one mastery tutoring lifts the average student  $\sim 2$  standard deviations (SD) above a classroom — does not survive contact with modern meta-analysis. The Nickow–Oreopoulos–Quan review of PreK-12 RCTs finds a pooled human-tutoring effect of  $\sim 0.37$  SD in the working paper (**0.29 SD** in the published version), with **no study reaching two sigma** [4][5]. AI tutoring should be benchmarked against  $\sim 0.3$  SD, not 2.0.

### BOTTOM LINE:

The best-evidenced AI tutoring raises a *weak* tutor toward an *average* one — a real, equity-relevant effect of roughly **+4 to +9 pp mastery at  $\sim \$20$ /tutor/year** [1] — not an autonomous machine delivering two sigma. The market is pricing the hype curve; the evidence curve says efficacy is real, bounded ( $\sim 0.3$  SD benchmark), and most valuable where human quality is lowest.

## 1. Context and Scope

AI tutoring sits at the intersection of two Scaffold coverage domains: *AI tutoring and LLMs in teaching and learning* and *efficacy research and the science of learning*, with a third lens — *edtech business models and unit economics* — governing whether any of it scales profitably.

The topic is urgent because the gap between marketing and measured efficacy has rarely been wider. Vendors invoke Bloom's "two sigma" as if delivered; valuations in 2025–26 priced a personalization revolution; and at least one heavily-marketed AI-first school chain (Alpha School / "2 Hour Learning") claims top-1-2% national percentiles on the strength of *internal, unaudited* metrics [10]. Meanwhile, the first *randomized* studies of LLM tutoring have only just landed, and they tell a more disciplined story.

**In scope:** what randomized and quasi-experimental evidence shows about learning effects; the distinction between assistive (human-in-the-loop) and autonomous tutoring; unit economics and the path from engagement to outcomes; adoption and market sizing; risks (over-reliance, equity,

measurement). **Out of scope:** general-purpose chatbot use, content-generation tools for teachers, and credentialing platforms except where they bear on the efficacy question. The **system boundary** for "efficacy" is a learning gain measured on an instrument *not* controlled by the vendor — the single hardest bar in this field, and the one most claims fail to clear.

## 2. Technology Landscape and State of the Art

### 2.1 The three architectures

AI tutoring is not one thing. It spans three architectures with very different evidence profiles, costs, and failure modes.

1. **Assistive / human-in-the-loop** — AI coaches a live human tutor (Tutor CoPilot). The human remains the safety and pedagogy backstop; the AI supplies expert-like moves in real time. Highest-quality causal evidence; lowest safety risk.
2. **Autonomous, pedagogically-tuned** — a model fine-tuned for teaching (Google's LearnLM, the tutoring modes of Khanmigo built on GPT-class models) interacts directly with the learner, ideally Socratic rather than answer-giving. Strong session-level results; thin durable-gain evidence.
3. **Autonomous, generic** — a general chatbot (ChatGPT, Gemini, Claude) used as a de-facto tutor. Massive adoption, near-zero pedagogical guardrails, and the architecture most associated with *over-reliance* and answer-offloading — the same dynamic that gutted Chegg's business (see §Market & Equity Implications).

### 2.2 What the evidence shows by architecture

Architecture	Lead evidence	Design / sample	Headline result	Outcome type	Maturity
Assistive (human-in-loop)	Tutor CoPilot, Stanford [1][2]	RCT, ≈900 tutors / ≈1,800 K-12 students	+4 pp mastery overall; +9 pp for lowest-rated tutors ( $p < 0.01$ )	Topic mastery (graded)	Demonstrated
Autonomously, tuned	LearnLM, Google DeepMind [3]	RCT, 165 students, 5 UK schools, 7 wks	93.0% vs 91.2% 2nd-attempt accuracy; +5.5 pp transfer (96% approved msgs)	Session-level task success	Early / promising
Autonomously, tuned	Khanmigo (GPT-4 class) [6]	Mixed-methods, 69 undergrads + paper arm	Significant gains in all arms; **no significant difference between arms**	Concept test	Inconclusive on causal lift
Autonomously, generic	None (no controlled efficacy trial)	—	Adoption high; efficacy unproven; over-reliance risk	—	Unproven / risky
Human 1:1 (benchmark)	Nickow–Oreopoulos–Quan [4][5]	Meta-analysis of PreK-12 RCTs	~0.37 SD (WP) / 0.29 SD (pub); none reach $2\sigma$	Standardized achievement	Established

The pattern is consistent across the credible studies: **AI tutoring works best as an amplifier of human pedagogy, and its measured effects are real but modest** — in the ballpark of, not multiples beyond, good human tutoring. Where the AI is left fully autonomous and rigorously tested (Khanmigo's controlled arm), it has so far failed to separate from cheaper comparison conditions on learning *outcomes*, even as students rate the experience highly [6].

### 2.3 The two-sigma anchor — and why it misleads

Bloom (1984) reported tutored students scoring  $\sim 2$  SD above classroom peers ( $2.18\sigma$  for geometry/algebra;  $2.30\sigma$  for adult Korean) with mastery learning alone at  $\sim 1\sigma$  [7][8]. Two structural problems undercut using this as the AI benchmark. First, the tutoring arms held students to a **90% mastery threshold vs 80%** in the classroom arm — the goal level was not held constant, which alone can manufacture much of the gap (van Lehn's critique) [8]. Second, no modern randomized tutoring study reproduces it: the pooled effect is  $\sim 0.3$  SD [4][5]. A defensible expectation for *autonomous AI* tutoring is therefore at or below the human-tutoring frontier of  $\sim 0.3$  SD until proven otherwise — roughly one-sixth of the figure marketers cite.

#### THE TAKE:

The two-sigma number is functioning as a *category error* in the AI-tutoring market. It described a tightly-controlled lab condition (held to a higher mastery bar) and has never been reproduced in the field at even half its magnitude. Treating it as the addressable prize lets vendors frame a  $\sim 0.3$  SD product as a  $\sim 2.0$  SD miss-in-progress. The correct mental model is not "close the two-sigma gap" but "**deliver  $\sim 0.3$  SD reliably, at near-zero marginal cost, to the students human tutoring never reaches**" — a smaller per-student effect whose value is in *distribution*, not magnitude.

## 3. Techno-Economic Analysis

### 3.1 From engagement to outcomes — the cost model

The economics of AI tutoring hinge on a chain the marketing usually skips: **inference cost -> engagement -> completion -> measured learning gain -> willingness to pay**. Each link leaks. We model the per-learner unit economics of an autonomous tutoring product against the human-tutoring benchmark, with every input cited or labelled *Scaffold estimate*.

### 3.2 Cost Model and Assumptions

Parameter	Value	Unit	Basis / Source
Tutor CoPilot (assistive) cost	$\sim 20$	USD / tutor / year	Stanford RCT disclosure [1]
High-dosage human tutoring cost	$\sim 3,500$ – 4,300	USD / student / year	High-dosage program range, U.S. K-12 [9]; *Scaffold estimate of midpoint*
Human tutoring pooled effect	0.29–0.3 7	SD	Nickow et al. [4][5]
Autonomous AI inference cost (2026)	declining YoY	USD / session	DUOL gross margin +190 bps YoY on per-unit AI cost cuts [11]
DUOL gross margin (Q1 FY26)	73.0	%	Duolingo 8-K, Q1 FY26 [11]
Assumed engagement->completion rate	15–35	%	*Scaffold estimate* (MOOC-era completion norms, applied to self-serve AI tutoring)
AI in education market (2025)	$\sim 6.9$ – 7.0 5	USD bn	Mordor / Precedence [12][13]
AI tutors sub-market (2025->2030)	3.55 -> 6.45	USD bn	Market research, 2025 vintage [12]

### 3.3 Unit economics and the path to outcomes

The structural advantage of autonomous AI tutoring is brutal on cost: the marginal cost of a tutoring session is an inference call that is **falling year-over-year**, visible in Duolingo's gross margin expanding 190 bps to 73.0% in Q1 FY26 "driven primarily by continued reductions in per-unit AI costs" [11]. Against high-dosage human tutoring at ~\$3,500–4,300 per student per year [9], an AI tutor that delivers even a fraction of the ~0.3 SD human effect at <1% of the cost is economically transformative — *if* the effect survives outside the lab.

The assistive model is the cleanest economic story today: Tutor CoPilot's ~\$20/tutor/year buys a measured +4 to +9 pp mastery effect [1]. Amortized across the ~20–30 students a tutor serves, the cost per student-equivalent learning gain is **on the order of \$1 (Scaffold estimate)** — two to three orders of magnitude below high-dosage human tutoring for a fraction of the effect, but a *real, measured* fraction. That is the most favorable cost-per-confirmed-outcome ratio in this report.

#### THE TAKE:

The unit-economics winner and the efficacy winner are, for now, the *same* architecture — assistive. Autonomous tutoring has the better long-run cost curve (inference, not labor) but has not closed the outcome-evidence gap; assistive tutoring has the proof *and* a cost low enough to be a rounding error on a tutoring budget. Until an autonomous product publishes a vendor-independent durable learning gain, the rational procurement bet is **AI that makes humans better, not AI that replaces them**. *Scaffold estimate*: the first autonomous product to clear that bar with a pre-registered, externally-scored RCT showing  $\geq 0.15$  SD will command a pricing premium the current crop cannot.

### 3.4 Sensitivity

The answer is dominated by four variables. We rank them by how much they move the conclusion (most-sensitive first):

Driver	Low case	High case	Effect on conclusion
Durable vs session-level gain	Session-only (current LearnLM/Khanmigo)	Externally-audited $\geq 0.2$ SD durable	Largest swing: flips autonomous AI from "interface" to "intervention"
Engagement / completion	15% complete	35%+ complete	Self-serve consumer AI tutoring lives or dies here; near-linear on realized ARPU
Inference cost trajectory	Flat	Continued YoY decline [11]	Sets gross margin ceiling; favors scaled platforms
Human-quality baseline	High (well-trained tutors)	Low (under-resourced)	Determines assistive uplift: +4 pp avg vs +9 pp for weak tutors [1]

The single highest-leverage uncertainty is the first: whether autonomous AI tutoring produces *durable, transferable* learning gains on instruments the vendor does not control. Every credible 2025–26 study measured something nearer the session than the semester.

## 4. Market and Demand Outlook

The addressable opportunity is large and growing fast, but the headline "AI in education" numbers blend tutoring with content tools, admin, and assessment. Reported sizings (2025 vintage): the global edtech market ~\$187–189 bn in 2025 [13][14]; **AI-in-education ~\$6.9–7.05 bn in 2025**, projected to ~\$41 bn by 2030 (CAGR ~43%) [12][13]; and the narrower **AI-tutors sub-segment ~\$3.55 bn in 2025 -> ~\$6.45 bn by 2030** [12]. Treat the high-CAGR figures as directional — they assume efficacy and willingness-to-pay that the evidence has not yet confirmed.

Adoption is running well ahead of proof. Duolingo reached **56.5 M daily active users (+21% YoY)**

and **12.5 M paid subscribers** in Q1 FY26, with subscription revenue of **\$250.9 M (+31% YoY)** and a target of 100 M DAU by 2028 [11]. Coursera posted **FY2025 revenue of \$757.5 M (+9%)**, tripled its generative-AI catalog to 925+ courses past 10 M enrollments, signed an Anthropic content partnership, and agreed an all-stock merger with Udemy [15]. Pearson reported **2025 sales of £3,577 m (+4% underlying)**, £614 m adjusted profit, 20% Q4 growth in Virtual Learning, and an IBM AI-tutoring partnership [16]. The demand signal is unambiguous; the *efficacy* signal is the lagging variable.

#### 4.1 Scenarios to 2030

We frame three scenarios for AI-tutoring efficacy-and-adoption, with the pivotal sensitivity attached to each.

**Scenario A — "Amplifier" (base, ~50% Scaffold estimate).** Assistive AI is the proven workhorse; autonomous AI remains a strong *engagement* layer with session-level but not durable-gain evidence. AI-tutors sub-market tracks the ~\$6.45 bn-by-2030 path [12]. Winners are platforms that pair scale with measurement (DUOL, GOOGL infrastructure); pure-play homework substitutes keep eroding. *Pivot: engagement monetization holds; no autonomous efficacy breakthrough.*

**Scenario B — "Intervention" (upside, ~25% Scaffold estimate).** One or more autonomous products publishes a pre-registered, vendor-independent RCT showing  $\geq 0.2$  SD durable gains. Procurement budgets shift from human tutoring toward AI-augmented models; the AI-in-education market overshoots the ~\$41 bn-by-2030 line [13]. *Pivot: durable-gain evidence materializes; the cost curve does the rest.*

**Scenario C — "Backlash" (downside, ~25% Scaffold estimate).** Over-reliance, surveillance, and unaudited-claims controversies (the Alpha School template [10]) trigger procurement caution and regulatory scrutiny; consumer novelty fades and completion rates disappoint. Growth decelerates toward the lower edtech CAGR (~10–11%) [14] and reprices the high-multiple names. *Pivot: measurement and trust fail before efficacy is proven.*

## 5. Feasibility, Scale-Up, and Risk

AI tutoring is *technically* feasible at scale today — inference is cheap and falling, and the assistive model is already deployed in live tutoring at thousands-of-tutor scale [1]. The binding constraints are not compute; they are **evidence, trust, and measurement**.

### 5.1 Risk Register

Risk	Likelihood	Impact	Mitigation
Over-reliance / answer-offloading (learning *displaced*, not aided)	High	High	Socratic/guardrailed design; the Chegg->ChatGPT collapse is the cautionary case [17][18]
Efficacy is session-level only, not durable	High	High	Pre-registered RCTs scored on external instruments; longitudinal follow-up
Equity cuts both ways	Medium	High	AI most helps weak-tutor settings (+9 pp [1]) — but only where access, devices, and supervision exist
Measurement capture (vendor-controlled metrics)	High	High	Independent evaluation; reject internal-only claims (Alpha School precedent [10])
Safety / harmful or wrong content	Low-Med	High	Human-in-loop review (LearnLM: 0 harmful, 0.1% factual errors [3])

Surveillance / privacy backlash	Medium	Medium	Minimize monitoring; transparent data practices [10]
Inference-cost or model-access shock	Low	Medium	Multi-model sourcing; the platform owners (GOOGL) hold leverage

The go/no-go read: **assistive AI tutoring is a "go" on current evidence; autonomous AI tutoring is "conditional"** — feasible and cheap, but not yet proven to change durable outcomes, and carrying real over-reliance and measurement risk if deployed without guardrails.

## 6. Market and Equity Implications

This section names exposed public companies and the directional read tied to the report's thesis — *assistive AI is the proven, low-cost winner; autonomous efficacy is unproven; engagement is being priced ahead of outcomes; and generic-chatbot substitution is a destroyer of narrow homework businesses*. It is **not** investment advice (see Disclosures). Tickers verified; figures cited to public sources.

Company (Ticker)	Exposure	Reasoning (tied to the thesis)	Horizon
Duolingo (DUOL, Nasdaq)	Positive / engagement-led	Best-monetized AI engagement layer; 56.5 M DAU, \$250.9 M sub revenue +31% YoY, 73.0% gross margin on falling AI cost [11]. Strength is engagement+cost, not audited efficacy — rewarded in Scenario A, exposed in C.	12–24 mo
Coursera (COUR, NYSE)	Neutral-to-Positive	AI-course demand real (925+ courses, 10 M+ enrollments, FY25 rev \$757.5 M +9%) [15]; Udemy merger adds scale but integration + content-commoditization risk caps the upside.	12–24 mo
Chegg (CHGG, NYSE)	Negative	The clearest casualty of generic-chatbot substitution: rev –30% YoY (Q1'25), subs –31% then –43% by Q3, ~\$1 share, multiple restructurings [17][18]. Validates the over-reliance thesis as a *business* risk.	0–12 mo
Pearson (PSO, NYSE/LSX)	Neutral	Incumbent with distribution + AI partnerships (IBM); 2025 sales £3,577 m +4%, Virtual Learning +20% Q4 [16]. Defensible but low-growth; AI is a margin/retention story, not a re-rating.	12–36 mo
Alphabet (GOOGL, Nasdaq)	Positive (platform exposure)	Owns the most credible *autonomous* efficacy research (LearnLM RCT [3]) and the model/infrastructure layer many tutors run on. Exposure is optionality on Scenario B, not a pure-play.	24–36 mo

### THE TAKE:

The market is paying for *engagement* (DUOL's multiple) and punishing *substitution* (CHGG's collapse), but it has not yet repriced for *audited efficacy* — because no autonomous product has delivered it. The non-obvious implication: the first company to publish a pre-registered, vendor-independent durable learning gain doesn't just win a study — it redefines edtech procurement from "engagement spend" to "outcomes spend," and the incumbents best positioned to monetize that shift are the *platform* (GOOGL) and the *measurement-credible* operators, not the homework-substitution layer that generic chatbots already commoditized. *Scaffold estimate*: on current trajectories, that proof point is more likely 2027–2028 than 2026.

## 7. Outlook and Strategic Implications

AI tutoring's efficacy question has a defensible answer today, and it is neither the hype nor the dismissal. **The effect is real, modest (benchmark ~0.3 SD, not 2.0), and most valuable where human teaching quality is lowest** — an equity instrument more than a genius-maker. The assistive architecture has the proof and the economics; the autonomous architecture has the cost curve and the adoption but still owes the field a durable, externally-scored learning gain. Operators should buy *AI that makes humans better* now and treat *AI that replaces them* as a watchlist bet contingent on evidence. Investors should distinguish engagement multiples (priced) from outcomes evidence (not yet priced), and weight the over-reliance/measurement risks the Chegg and Alpha School cases have already made concrete.

### WHAT TO WATCH:

- **A pre-registered, vendor-independent RCT** on autonomous AI tutoring reporting durable, transfer-scored gains  $\geq 0.15$ – $0.2$  SD — the single signpost that flips Scenario A->B. None published as of June 2026. - **Duolingo's path to 100 M DAU by 2028** [11] and whether engagement converts to retained paid subs as AI novelty normalizes (Scenario A vs C). - **Independent scrutiny of internal-metric claims** (Alpha School / "2 Hour Learning" [10]) — a high-profile debunk would harden the Scenario-C backlash. - **Inference-cost and gross-margin trajectory** at the scaled platforms (DUOL's 73.0% and rising [11]) — the floor under autonomous unit economics.

## Disclosures & Disclaimer

This report is general commentary published for information purposes only. It is **not** investment advice, a recommendation, or a solicitation to buy or sell any security, and it does not account for the objectives or circumstances of any individual. Scaffold is a research publication, not a registered investment adviser or broker-dealer. Views are the publication's own analytical opinions, are subject to change, and may prove wrong. Markets involve risk of loss; past performance does not indicate future results. Readers should do their own research and consult a licensed financial professional before acting. The publication and/or its principals may hold positions in securities mentioned. Company facts and figures are drawn from public sources believed reliable but are not guaranteed. © Scaffold.

## Methodology and Assumptions

This report weighs *causal* evidence over correlational or marketing claims. The evidence hierarchy applied, strongest first: (1) randomized controlled trials with learning-outcome measures (Tutor CoPilot [1][2], LearnLM [3]); (2) systematic meta-analyses of RCTs for the human-tutoring benchmark (Nickow–Oreopoulos–Quan [4][5]); (3) mixed-methods controlled studies (Khanmigo [6]); (4) company financial disclosures and filings for adoption and unit economics (DUOL [11], COUR [15], CHGG [17][18], PSO [16]); (5) third-party market sizings, flagged as directional given divergent definitions [12][13][14].

The "efficacy" bar is a learning gain measured on an instrument the vendor does not control; internal-only metrics (e.g. Alpha School [10]) are reported as claims, not evidence. The two-sigma figure is treated as a historically-contested anchor, not a target [7][8]. Every quantitative claim is cited; figures derived by Scaffold are labelled "Scaffold estimate" with their inputs shown in the cost-model and sensitivity tables. Effect sizes are reported in their original units (percentage points for mastery/accuracy where the source uses them; SD for the meta-analytic benchmark); these are not interchangeable, and the report avoids converting between them. The conclusion would change most if

a pre-registered, vendor-independent RCT demonstrated durable, transfer-scored gains from *autonomous* AI tutoring – the signpost named throughout. Data vintage: financials Q1 FY2026 and FY2025; studies 2024–2025; market sizings 2025.

## References

1. Wang, R.E., Ribeiro, A.T., Robinson, C.D., Loeb, S., Demszky, D., et al. "Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise." Stanford / arXiv:2410.03017, 2024. <https://arxiv.org/abs/2410.03017>
2. Stanford SCALE / National Student Support Accelerator. "Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise" (study summary). 2024. <https://nssa.stanford.edu/studies/tutor-copilot-human-ai-approach-scaling-real-time-expertise>
3. Google DeepMind LearnLM team. "AI tutoring can safely and effectively support students: An exploratory RCT in UK classrooms." arXiv:2512.23633, 2025. <https://arxiv.org/html/2512.23633v1>
4. Nickow, A., Oreopoulos, P., Quan, V. "The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence." NBER Working Paper No. 27476, 2020. <https://www.nber.org/papers/w27476>
5. Nickow, A., Oreopoulos, P., Quan, V. "The Promise of Tutoring for PreK–12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence." *American Educational Research Journal*, 2024. <https://journals.sagepub.com/doi/10.3102/00028312231208687>
6. "Leveraging 'Khanmigo' Generative AI-Powered Tool for Personalized Tutoring to Learn Scientific Concepts." *Journal of Teaching and Learning*, 2025. <https://jtl.uwindsor.ca/index.php/jtl/article/view/10052>
7. Bloom, B.S. "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring." *Educational Researcher* 13(6), 1984. [https://en.wikipedia.org/wiki/Bloom's\\_2\\_sigma\\_problem](https://en.wikipedia.org/wiki/Bloom's_2_sigma_problem)
8. "Two-Sigma Tutoring: Separating Science Fiction from Science Fact." *Education Next*, summarizing van Lehn's critique of Bloom (1984). <https://www.educationnext.org/two-sigma-tutoring-separating-science-fiction-from-science-fact/>
9. FutureEd, Georgetown University. "Research Notes: Two Emerging Strategies for Using AI in Tutoring" (high-dosage tutoring cost context). <https://www.future-ed.org/research-notes-two-emerging-strategies-for-using-ai-in-tutoring/>
10. "Alpha School" (overview, claims, and independent-scrutiny summary). Wikipedia; with reference to WIRED (Oct 2025) and 404 Media (Feb 2026) investigations. [https://en.wikipedia.org/wiki/Alpha\\_School](https://en.wikipedia.org/wiki/Alpha_School)
11. Duolingo, Inc. Form 8-K, Q1 FY2026 results (revenue \$292.0 M +27%; 56.5 M DAU; 12.5 M paid subs; subscription revenue \$250.9 M +31%; gross margin 73.0%). U.S. SEC, 2026. <https://www.sec.gov/Archives/edgar/data/0001562088/000162828026029790/q1fy26duoling03-31x26share.htm>
12. Mordor Intelligence. "AI in Education Market Size & Industry Trends Report 2030" (AI tutors sub-segment \$3.55 bn 2025 -> \$6.45 bn 2030). 2025. <https://www.mordorintelligence.com/industry-reports/ai-in-education-market>
13. Precedence Research. "AI in Education Market Size to Surge USD 136.79 Bn by 2035" (2025 base ~\$7.05 bn). 2025. <https://www.precedenceresearch.com/ai-in-education-market>
14. Grand View Research. "Education Technology Market Size | Industry Report, 2033" (global edtech ~\$187 bn, 2025). 2025. <https://www.grandviewresearch.com/industry-analysis/education-technology-market>

15. Coursera, Inc. FY2025 results and AI-catalog / Anthropic partnership / Udemy merger disclosures (revenue \$757.5 M +9%; 925+ gen-AI courses; 10 M+ enrollments). 2025–2026.  
<https://www.stocktitan.net/news/COUR/>
16. Pearson plc. 2025 Preliminary Results (sales £3,577 m +4% underlying; adjusted profit £614 m; Virtual Learning +20% Q4; IBM partnership). 2026.  
<https://www.stocktitan.net/news/PSO/pearson-2025-preliminary-results-3joib741fl2v.html>
17. Chegg, Inc. Form 10-Q, Q3 FY2025 (subscription revenue –37% for nine months; subscribers –43% YoY). U.S. SEC, 2025.  
<https://www.sec.gov/Archives/edgar/data/0001364954/000136495425000117/chgg-20250930.htm>
18. Chegg, Inc. Form 8-K, Q1 FY2025 results (revenue \$121.4 M –30% YoY; 3.2 M subscribers –31% YoY; generative-AI substitution disclosure). U.S. SEC, 2025.  
<https://www.sec.gov/Archives/edgar/data/0001364954/000136495425000049/a9901-financialresultsq120.htm>

## About Scaffold

Scaffold publishes independent research on learning technology — AI tutoring and LLMs in teaching and learning, learning-management and courseware platforms, assessment and learning analytics, credentialing and alternative pathways, the science of learning and its efficacy evidence, and the funding, business models and unit economics of edtech. Reports are prepared for subscribers and are provided for information only; they do not constitute investment, legal, or educational advice. © Scaffold. All rights reserved.