

Scalable Personalization in Online Education: AI-Driven Implementation Systems vs. Traditional Coaching Models

Humberto Inciarte

Department of Research, Arquitectos de la Felicidad / NeuroFlow 30H™

ORCID: 0009-0006-0379-7328

Corresponding Author: hola@humbertoinciarte.com

May 2026

Abstract

The prevailing response among independent course creators to declining completion rates and rising student demand for personalization is to increase the volume of synchronous instructor presence — additional live sessions, expanded one-to-one coaching, and longer office hours. This paper argues that this response misidentifies the bottleneck. Drawing on four decades of empirical research, from Bloom’s (1984) two-sigma problem to recent randomized controlled trials of artificial-intelligence (AI) tutoring (Kestin et al., 2025; De Simone et al., 2025; Wang et al., 2024), the analysis demonstrates three convergent findings. First, the personalization–scale trade-off is structurally bounded by the human time budget; one-to-one delivery cannot be the answer at population scale (Baumol & Bowen, 1966). Second, the silent dropout — the student who disengages without ever requesting help — is the dominant failure mode of online education and is driven by social cost rather than personality (Karabenick & Knapp, 1988; Ryan & Pintrich, 1997). Third, well-engineered AI implementation systems produce learning effects statistically comparable to one-to-one human tutoring (Ma et al., 2014; VanLehn, 2011) at population sizes orders of magnitude beyond synchronous capacity. The paper proposes the CursoVivo framework — embedding AI-driven personalization within existing course structures to encode the creator’s professional judgment into a system that delivers individualized intervention twenty-four hours a day, without the creator’s real-time presence. The framework is presented as a model proposition, not as an empirical claim, with explicit acknowledgment of contemporary counter-evidence (Bastani et al., 2024) on the conditional value of generative AI in educational contexts.

Keywords: scalable coaching, AI personalization, online education AI, course completion, automated coaching, e-learning personalization, implementation systems, CursoVivo, *aprendizaje personalizado escalable, sistemas de implementación con inteligencia artificial*

1. Introduction

1.1 The Prevailing Narrative

Across the global creator economy, the consensus prescription for failing online courses is more of the creator. When completion rates fall, when students complain that material feels generic, when prospects ask whether the same content can be obtained through ChatGPT, the recommendation funneled to course creators by platforms, consultants, and peers is consistent: add more live sessions, open more one-to-one coaching slots, increase synchronous touchpoints, and demonstrate human availability. The Billion Dollar Boy (2024) creator-economy study found that fifty-two percent of creators report career burnout and thirty-seven percent have considered abandoning the field — a pattern consistent with a labor model that responds to declining outcomes by demanding more of the creator’s bounded time.

This narrative is intuitive. It rests on a defensible premise (personalized attention improves outcomes) and an undefended inference (therefore, more of the instructor’s personal attention is the path to better outcomes). The premise is supported by Bloom’s (1984) classic finding that one-to-one mastery tutoring produces learning gains roughly two standard deviations above conventional group instruction. The inference, however, ignores the fact that Bloom himself concluded the method was “too costly for most societies to bear on a large scale” and posed the explicit research challenge of finding scalable group methods that approximate the tutoring effect (Bloom, 1984, p. 4). For four decades, that challenge stood largely unsolved.

1.2 The Problem

The cost of the prevailing prescription is structural rather than motivational. A course creator with two hundred enrolled students who attempts to provide weekly one-hour personal contact would commit ten thousand four hundred hours annually to that single function — a quantity that exceeds the total billable capacity of any individual practitioner. The mathematics of synchronous delivery sets a hard ceiling on personalization-per-student. As that ceiling is approached, two failure modes emerge in parallel: creator burnout and student disengagement. Reich and RUIPÉREZ-VALIENTE’S (2019) analysis of all Massachusetts Institute of Technology and Harvard University massive open online courses (MOOCs) on the edX platform between 2012 and 2018 found that fifty-two percent of registrants never began coursework, and that completion among those who started fell from approximately six percent in 2014 to 3.13 percent by 2017–2018. Even among paying “verified” learners, completion declined from fifty-six to forty-six percent over the same period. HONE AND EL SAID (2016) found that instructor interaction quality alone explained seventy-nine percent of the variance in MOOC retention — a finding that simultaneously vindicates the importance of instructor presence and indicts the impossibility of supplying it at scale through synchronous delivery.

1.3 Research Question and Thesis

This paper examines whether the prevailing response — *more presence of the instructor* — holds up under empirical scrutiny, and proposes an alternative framework — the CursoVivo implementation model — that addresses the structural constraints identified in the literature. The central research question is: *Can the personalization effect documented in one-to-one tutoring be reproduced at population scale through AI-driven implementation systems, without compromising educational outcomes?* The working thesis advanced

is that the time ceiling on synchronous delivery is not a problem to be solved by working harder within the synchronous paradigm; it is a structural limit that requires the encoding of professional judgment into algorithmically-scalable systems. The paper is organized as follows: Section 2 reviews the empirical literature on completion rates, help-seeking behavior, coaching effectiveness, and AI tutoring outcomes. Section 3 presents the analytic case for AI-driven personalization and articulates the CursoVivo framework as a model proposition, alongside a candid treatment of contrary evidence. Section 4 concludes with limitations and a research agenda.

2. Literature Review

2.0 Methodological Note

This review synthesizes peer-reviewed empirical studies, randomized controlled trials, institutional reports, and meta-analyses published between 1984 and 2025, sourced from PubMed, PsycINFO, Google Scholar, arXiv, the World Bank Open Knowledge Repository, and the SAGE and Wiley journal portfolios. Inclusion criteria prioritized studies with measurable learning, retention, or behavioral outcomes in adult or post-secondary online education contexts, supplemented by foundational K-12 and laboratory studies where the underlying mechanism (e.g., help-seeking, retrieval practice) is established as domain-general. Industry sources (Khan Academy, Ruzuku, Billion Dollar Boy) were included only to provide contemporary contextual benchmarks; load-bearing claims rely on peer-reviewed citations.

2.1 The Documented Failure of Scale Without Personalization

The empirical record on standard MOOC completion is consistent across institutions, languages, and time. Reich and Ruipérez-Valiente (2019) found completion rates of approximately three percent in MIT-Harvard offerings. Onah, Sinclair, and Boyatt (2014) reported MOOC completion rates “typically below 13%” across the field, with the supported instructional mode (incorporating tutored laboratory components) outperforming the unsupported standard mode in their University of Warwick deployment. Khalil and Ebner’s (2014) literature synthesis identified five recurring drivers of MOOC dropout: lack of time, lack of motivation, *isolation and lack of interactivity*, insufficient prerequisite skills, and hidden costs. The third driver is decisive for the present argument: isolation is not a side-effect of poor course design but the direct consequence of an architecture in which one instructor faces an unbounded student population through asynchronous video.

Hone and El Said (2016) administered a structured retention survey to three hundred seventy-nine MOOC participants in Cairo and identified instructor interaction and course content quality as the two strongest predictors of retention, jointly explaining seventy-nine percent of the variance in their regression model. The finding is a paradox at the heart of online education: the variable most strongly correlated with completion is the variable that synchronous human delivery cannot supply at population scale.

2.2 The Silent Dropout: Help-Seeking Asymmetries

A central but under-recognized failure mode of online education is the structural unwillingness of struggling students to request help. Karabenick and Knapp’s (1988) controlled experiment provided one of the

field's most striking demonstrations: when students were offered help from a computer interface (private), eighty-six percent sought assistance; when the same help was offered from a person (public), only thirty-six percent did. The 2.4-fold difference indicates that help-avoidance is not a personality trait but a response to perceived social cost. Ryan and Pintrich (1997) extended this finding by showing that help-avoidance is most pronounced in students with the lowest self-efficacy and competence — that is, the students who would benefit most from intervention are those least likely to request it.

This pattern is reinforced by participation-inequality effects in online communities. Nielsen's (2006) widely-replicated 90-9-1 rule documents that approximately ninety percent of online community members never contribute, nine percent contribute occasionally, and one percent generate the majority of content. Sinha, Jermann, and Dillenbourg (2014) found that course-discussion forums in MOOC environments engage only approximately five percent of enrolled learners, leaving ninety-five percent of struggling students invisible to forum-mediated support. In intelligent-tutoring-system contexts, more than seventy percent of users have been shown to exhibit help-seeking deviations, including help-avoidance and over-reliance on hints (Aleven et al., reviewed in narrative literature on online-learning behaviors). The cumulative implication is that the standard "open-door" support model — wait for the student to raise a hand — is not merely suboptimal; it systematically excludes the population most in need of intervention.

2.3 Coaching Effectiveness and the Format-Equivalence Finding

A separate body of evidence addresses whether coaching, defined here as structured one-to-one developmental interaction, produces measurable outcomes — and, critically, whether synchronous delivery format moderates those outcomes. Theeboom, Beersma, and van Vianen (2014) conducted a meta-analysis of organizational coaching effects across performance, well-being, coping, work attitudes, and goal-directed self-regulation, finding statistically significant positive effects in all five domains (Hedges' $g = 0.43-0.74$). Notably, the number of coaching sessions did not significantly moderate effectiveness — the dose-response curve was flat. Jones, Woods, and Guillaume (2016) replicated the central finding (overall $\delta = 0.36$; affective $\delta = 0.51$; individual results $\delta = 1.24$) and added a critical moderator analysis: face-to-face delivery did not differ significantly from blended or e-coaching formats on any outcome measure. The number of sessions and intervention longevity also failed to moderate outcomes.

These two meta-analyses constitute strong evidence for two propositions that bear directly on the present argument. First, coaching as a modality produces real, replicable effects. Second, the *delivery channel* of coaching — and the *quantity* of synchronous hours — is not the active ingredient. What appears to drive outcomes is the structured, personalized application of professional judgment to the recipient's specific context. This decouples the intuitive bundle "coaching = real-time human presence" and creates conceptual room for the proposition examined in Section 3: that the structured application of judgment can be encoded into systems that operate beyond the human time budget.

2.4 AI-Driven Personalization at Scale

Four decades of research on intelligent tutoring systems (ITS), and a recent generation of large-language-model-based tutoring studies, provide convergent evidence that algorithmic personalization can approximate or exceed the effect sizes of small-group human tutoring. VanLehn's (2011) meta-review of twenty-eight

controlled studies found that step-based ITS produced effect sizes ($d \approx 0.76$) statistically comparable to those of one-to-one human tutoring ($d \approx 0.79$). Kulik and Fletcher (2016), reviewing fifty controlled evaluations, reported a median ITS effect of 0.66 standard deviations on test scores — corresponding to a movement from the fiftieth to the seventy-fifth percentile. Ma, Adesope, Nesbit, and Liu (2014) found ITS outperformed teacher-led large-group instruction ($g = 0.42$) and conventional computer-based instruction ($g = 0.57$), with no statistically significant difference relative to small-group human tutoring.

Recent randomized controlled trials with generative-AI-based tutors extend these findings to the contemporary technological context. Kestin, Miller, Klales, Milbourne, and Ponti (2025) randomized one hundred ninety-four Harvard Physical Sciences 2 students to either an instructor-led active-learning classroom or to a custom GPT-4-based tutor with explicit pedagogical scaffolding. The AI-tutored group achieved more than double the learning gains in less time, with higher self-reported engagement and motivation. De Simone et al. (2025), in a World Bank-published trial of eight hundred Edo State (Nigeria) secondary students, documented gains of 0.31 standard deviations after six weeks of after-school AI-tutored sessions — equivalent to roughly 1.5 to 2 years of business-as-usual schooling, and ranking the intervention among the most cost-effective educational programs ever documented. Wang, Ribeiro, Robinson, Loeb, and Demszky’s (2024) Tutor CoPilot trial — nine hundred tutors and one thousand eight hundred Title I students — showed that pairing human tutors with an AI assistant raised topic mastery by four percentage points overall and nine percentage points among students assigned to the lowest-rated tutors, at a cost of approximately twenty dollars per tutor annually.

2.5 Research Gap

The literature thus documents (a) the structural failure of unscaled MOOCs, (b) the silent-dropout problem rooted in help-seeking asymmetries, (c) the format-independence of coaching outcomes, and (d) the empirical viability of AI-driven personalization at scale. What it has not yet articulated is a unified framework for course creators in non-academic, independent-creator settings — the population that operates outside the institutional resources of MIT, Harvard, or Stanford — by which professional judgment can be systematically encoded into a deployable system. The CursoVivo model proposed in Section 3 addresses this gap.

3. Analysis and Discussion

3.1 The Central Empirical Finding: Personalization at Scale Is No Longer Theoretical

The 2014 Ma et al. meta-analytic finding — that intelligent tutoring systems produce no statistically significant difference in learning outcomes relative to small-group human tutoring — should be regarded as a quietly transformative result for the economics of education. For the four decades following Bloom’s (1984) original formulation, the dominant assumption in educational economics was that personalization and scale stood in inverse relation. The 2014 finding, corroborated by VanLehn (2011), Kulik and Fletcher (2016), and now reinforced by Kestin et al. (2025) and De Simone et al. (2025) in the generative-AI era, breaks the inverse relation. The implication for course creators is direct: the educational outcome historically produced by a tutor sitting beside one student can now be approximated, in measurable effect-size terms, by a

system serving thousands of students concurrently — provided the system encodes the relevant professional judgment with adequate fidelity.

3.2 Quantifying the Time Ceiling: An Economic Restatement

The structural argument can be expressed in straightforward labor economics. A course creator serving two hundred enrolled students who attempts to provide one hour of weekly synchronous personal contact commits two hundred hours per week — a quantity exceeding the total weekly working time of any single individual. Reduced to a more conservative monthly check-in of one hour per student, the figure becomes two thousand four hundred hours annually for support delivery alone, before content creation, marketing, administration, or sales. Baumol and Bowen's (1966) formulation of the cost disease in labor-intensive services applies directly: in any service whose unit of production is the practitioner's time, productivity gains can come only from technology that substitutes for the practitioner's *judgment*, not technology that merely reproduces or distributes it. Pre-recorded video, the dominant technology of the contemporary creator economy, distributes content but does not substitute for judgment — it scales information without scaling guidance. The result is the asymmetry documented in Section 2.1: rising enrollment with falling completion.

3.3 The Silent Dropout as Structural, Not Motivational

The conventional account of student dropout attributes the failure to learner characteristics — insufficient discipline, poor time management, lack of motivation. The Karabenick and Knapp (1988) finding that help-seeking more than doubles when the helper is a computer rather than a person reframes the problem. The structural variable is not motivation but the social cost of disclosure of difficulty. Ryan and Pintrich (1997) showed that this cost is highest precisely for the population whose competence is lowest. The combination produces a self-reinforcing exclusion mechanism: the students most likely to need intervention are most likely to remain silent, and the intervention is offered through a channel (forums, office hours, public Q&A) whose structure imposes maximum social cost. AI-mediated implementation systems carry the structural property identified in 1988 and recently reinforced by Wang et al. (2024): they remove the public dimension of help-seeking. The student who would not raise a hand in a live session, would not message a peer in a forum, and would not email the instructor will, in laboratory and field studies alike, type the question into a private interface. This is not a matter of preference. It is a documented behavioral asymmetry.

3.4 Counter-Evidence and Boundary Conditions

Honest assessment of the personalization-at-scale literature requires acknowledging three classes of limiting evidence. First, Bloom's (1984) original two-standard-deviation figure has been substantially revised downward by subsequent replication. The mastery-threshold differential between Bloom's tutored and conventional groups inflated the apparent effect; Ma et al.'s (2014) more recent estimates place ITS effects in the 0.4–0.7 standard deviation range. The CursoVivo model, accordingly, must be calibrated to this range, not to Bloom's original figure. Second, Steenbergen-Hu and Cooper (2014) found ITS effects at the post-secondary level ($g \approx 0.35$) lower than at K-12 levels, suggesting domain and population moderation effects that any specific deployment must address empirically.

Third, and most consequentially, Bastani, Bastani, Sungu, Ge, Kabakçı, and Mariman (2024) reported a randomized trial in which Turkish high-school students given unrestricted access to ChatGPT for mathematics preparation underperformed the control condition by approximately seventeen percent on the final examination. Critically, the harm was eliminated when access was restricted to a “GPT Tutor” version designed with pedagogical guardrails — though that version produced no measurable gains either. The Bastani finding is the most important contemporary boundary condition for any AI-in-education claim. It indicates that the educational effect of AI is not a property of model access but of *delivery design*: the same underlying language model produces a learning gain (Kestin et al., 2025) or a learning loss (Bastani et al., 2024) depending entirely on how it is scaffolded. This boundary condition is not a refutation of AI personalization; it is a specification of its conditions of effectiveness.

3.5 Proposed Framework: The CursoVivo Implementation Model

The CursoVivo implementation model proposes embedding AI-driven personalization within existing course structures rather than offering it as a separate platform. The framework treats the course creator’s accumulated professional judgment — the case-specific decisions, methodological emphases, sequencing logic, and corrective heuristics that constitute the creator’s expertise — as the primary asset to be scaled. The model proposes six functional components, derived from the convergent literature reviewed:

1. *Adaptive weekly planning* — operationalizing the implementation-intentions effect documented by Gollwitzer and Sheeran (2006), in which if-then specification produces approximately threefold gains in goal completion ($d = 0.65$ across ninety-four studies).
2. *Stateful follow-up with memory* — addressing the silent-dropout asymmetry by reaching the student before the student must raise a hand, consistent with Karabenick and Knapp (1988).
3. *Deliverable production support* — substituting active retrieval and application for passive consumption, consistent with Roediger and Karpicke (2006) on testing-effect gains in long-term retention.
4. *Progress dashboards* — providing the metacognitive monitoring associated with self-regulated learning gains.
5. *Daily next-action specification* — addressing the intention–behavior gap that drives MOOC dropout (Khalil & Ebner, 2014).
6. *Method-specific AI tutoring* — instantiating the scaffolded, pedagogically-prompted AI demonstrated effective by Kestin et al. (2025) and Wang et al. (2024), as opposed to the unrestricted access shown harmful by Bastani et al. (2024).

The framework is presented as a model proposition, not as an empirical claim. Its hypothesized effect sizes are bounded by the existing ITS literature (0.4–0.7 standard deviations on completion-correlated outcomes); validation in independent-creator-economy populations is identified in Section 4 as a research priority.

3.6 Practical Implications

Three implications for course creators follow from the analysis. First, the bottleneck in declining completion rates is not insufficient marketing, insufficient content production, or insufficient creator presence; it is the absence of personalization architecture. Adding live sessions to an architecture that does not reach the silent ninety percent treats a structural problem with a synchronous solution. Second, the appropriate target of automation is not the creator's content (which is the creator's competitive moat) but the creator's *judgment under common student conditions* — the recurring patterns of student difficulty for which the creator has, through experience, developed reliable corrective responses. Third, the choice between human and AI delivery is not a binary. The Tutor CoPilot results (Wang et al., 2024) indicate that the highest-leverage configuration combines human creator presence (where it is irreplaceable) with AI delivery of the creator's judgment (where it is the bottleneck). The creator continues to host live sessions for community and accountability; the AI handles the per-student personalization that no human schedule can absorb.

4. Conclusions

4.1 Summary of Findings

The empirical record reviewed in this paper supports four propositions. First, the failure mode of contemporary online education is structural rather than motivational; standard MOOC completion rates near three percent (Reich & Ruipérez-Valiente, 2019) reflect an architecture without personalization, not a population without will. Second, the dominant failure pattern within that architecture is the silent dropout, driven by help-seeking asymmetries that more than double in private versus public settings (Karabenick & Knapp, 1988). Third, the coaching literature shows that the active ingredient of effective developmental intervention is the structured application of professional judgment, not the synchronous delivery format; format itself does not moderate outcome (Jones et al., 2016; Theeboom et al., 2014). Fourth, well-engineered AI tutoring systems produce learning effects statistically comparable to small-group human tutoring (Ma et al., 2014; VanLehn, 2011) and, in recent generative-AI trials, more than double the gains of expert-led active-learning classrooms (Kestin et al., 2025).

The synthetic conclusion — that professional judgment, once encoded with adequate fidelity, becomes amenable to parallelization beyond the bounded capacity of synchronous one-to-one delivery — represents a structural alternative to the prevailing prescription of *more presence*. The CursoVivo framework operationalizes this conclusion as a model proposition: the encoding of the creator's judgment into a system that delivers personalized intervention to populations exceeding human time-budget capacity, without the creator's continuous real-time presence. The framework does not replace human connection; it creates connection where current architectures produce silence.

4.2 Limitations

This review is subject to the selection bias inherent in narrative literature reviews. The CursoVivo framework has not yet been validated through independent randomized controlled trials in independent-creator-economy populations; the effect-size estimates inherited from the ITS and AI-tutoring literatures may not generalize

without modification to the heterogeneous, short-form, voluntarily-enrolled populations characteristic of paid online courses outside formal educational institutions. The Bastani et al. (2024) boundary condition is acknowledged as a serious limit on naïve generalization from positive AI-tutoring trials. Several studies cited (Cui et al., 2019/2020; the Khan Academy 2024–25 report) involve vendor co-authorship or institutional reporting; corroboration with fully independent peer-reviewed work has been preferred where available. The review has not addressed mental-health, motivational, or sociocultural variables in detail; their interaction with implementation-architecture variables represents a substantial unresolved research domain.

4.3 Future Research Directions

Three lines of investigation follow from the present analysis. First, randomized controlled trials of AI-driven implementation systems in independent-creator-economy contexts — pre-registered, with course-creator-specific judgment encoded as the experimental manipulation — are required to establish whether the ITS effect sizes documented in formal-education contexts replicate in voluntary, paid, adult populations. Second, the interaction between human creator presence (community, accountability, identity) and AI delivery (per-student personalization, twenty-four-hour availability) merits direct empirical comparison; the Tutor CoPilot (Wang et al., 2024) hybrid configuration provides a credible template. Third, the present line of work intersects with ongoing research by the author on AI-mediated implementation in adjacent service domains, including small-business operations and contractor sales infrastructure; cross-domain integration of the encoded-judgment principle is a promising direction for the next phase of this program.

Resumen en Español

La respuesta predominante de los creadores independientes de cursos online ante la caída de las tasas de completación y la creciente demanda de personalización por parte de los estudiantes consiste en aumentar el volumen de presencia sincrónica del instructor — más sesiones en vivo, más coaching uno-a-uno, más horas de oficina. Este artículo argumenta que dicha respuesta identifica erróneamente el cuello de botella. A partir de cuatro décadas de evidencia empírica, desde el problema de las dos sigmas de Bloom (1984) hasta los recientes ensayos controlados aleatorizados sobre tutorización con inteligencia artificial (Kestin et al., 2025; De Simone et al., 2025; Wang et al., 2024), el análisis demuestra tres hallazgos convergentes. Primero, el equilibrio entre personalización y escala está estructuralmente limitado por el presupuesto temporal humano; la entrega uno-a-uno no puede ser la respuesta a escala poblacional (Baumol & Bowen, 1966). Segundo, el abandono silencioso — el alumno que se desconecta sin solicitar nunca ayuda — es el modo de fracaso dominante de la educación online y obedece a un costo social, no a un rasgo de personalidad (Karabenick & Knapp, 1988; Ryan & Pintrich, 1997). Tercero, los sistemas de implementación con IA bien diseñados producen efectos de aprendizaje estadísticamente comparables a la tutoría humana uno-a-uno (Ma et al., 2014; VanLehn, 2011) en poblaciones varios órdenes de magnitud superiores a la capacidad sincrónica. El artículo propone el marco CursoVivo — la incrustación de personalización dirigida por IA dentro de las estructuras de cursos existentes para codificar el criterio profesional del creador en un sistema que entrega intervención individualizada veinticuatro horas al día, sin la presencia en tiempo real del creador. El marco se presenta como una propuesta de modelo, no como una afirmación empírica, con reconocimiento explícito

de la evidencia contraria contemporánea (Bastani et al., 2024).

Palabras clave: coaching escalable, personalización con inteligencia artificial, IA en educación online, completación de cursos, coaching automatizado, personalización en e-learning, sistemas de implementación, CursoVivo

References

- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). *Generative AI can harm learning* (Wharton School Research Paper). SSRN. <https://doi.org/10.2139/ssrn.4895486>
- Baumol, W. J., & Bowen, W. G. (1966). *Performing arts: The economic dilemma*. Twentieth Century Fund.
- Billion Dollar Boy. (2024). *Burnout emerges as a barrier to growth in the creator economy*. <https://www.billiondollarboy.com/news/over-half-of-creators-face-burnout/>
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16. <https://doi.org/10.3102/0013189X013006004>
- Cui, W., Lynch, R., Smith, T., Tong, R., Yarnall, L., Shear, L., & Feng, M. (2020). When adaptive learning is effective learning: Comparison of an adaptive learning system to teacher-led instruction. *Interactive Learning Environments*, 31(2), 793–803. <https://doi.org/10.1080/10494820.2020.1808794>
- De Simone, M. E., Tiberti, F., Barron Rodriguez, M., Manolio, F., Mosuro, W., & Dikoru, E. J. (2025). *From chalkboards to chatbots: Evaluating the impact of generative AI on learning outcomes in Nigeria* (Policy Research Working Paper). World Bank. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099548105192529324>
- Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, 38, 69–119. [https://doi.org/10.1016/S0065-2601\(06\)38002-1](https://doi.org/10.1016/S0065-2601(06)38002-1)
- Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education*, 98, 157–168. <https://doi.org/10.1016/j.compedu.2016.03.016>
- Jones, R. J., Woods, S. A., & Guillaume, Y. R. F. (2016). The effectiveness of workplace coaching: A meta-analysis of learning and performance outcomes from coaching. *Journal of Occupational and Organizational Psychology*, 89(2), 249–277. <https://doi.org/10.1111/joop.12119>
- Karabenick, S. A., & Knapp, J. R. (1988). Effects of computer privacy on help-seeking. *Journal of Applied Social Psychology*, 18(6), 461–472. <https://doi.org/10.1111/j.1559-1816.1988.tb00029.x>
- Karabenick, S. A., & Knapp, J. R. (1991). Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. *Journal of Educational Psychology*, 83(2), 221–230. <https://doi.org/10.1037/0022-0663.83.2.221>

- Kestin, G., Miller, K., Klales, 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. <https://doi.org/10.1038/s41598-025-97652-6>
- Khalil, H., & Ebner, M. (2014). MOOCs completion rates and possible methods to improve retention — A literature review. In *Proceedings of EdMedia 2014* (pp. 1305–1313). Association for the Advancement of Computing in Education. <https://www.learntechlib.org/primary/p/147656/>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918. <https://doi.org/10.1037/a0037123>
- Nielsen, J. (2006). *Participation inequality: The 90-9-1 rule for social features*. Nielsen Norman Group. <https://www.nngroup.com/articles/participation-inequality/>
- Onah, D. F. O., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: Behavioural patterns. In *EDULEARN14 Proceedings* (pp. 5825–5834). IATED. <https://wrap.warwick.ac.uk/65543/>
- Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, 363(6423), 130–131. <https://doi.org/10.1126/science.aav7958>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Ryan, A. M., & Pintrich, P. R. (1997). “Should I ask for help?” The role of motivation and attitudes in adolescents’ help seeking in math class. *Journal of Educational Psychology*, 89(2), 329–341. <https://doi.org/10.1037/0022-0663.89.2.329>
- Sinha, T., Jermann, P., Li, N., & Dillenbourg, P. (2014). *Capturing “attrition intensifying” structural traits from didactic interaction sequences of MOOC learners*. arXiv:1409.5887. <https://arxiv.org/abs/1409.5887>
- Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students’ academic learning. *Journal of Educational Psychology*, 106(2), 331–347. <https://doi.org/10.1037/a0034752>
- Theeboom, T., Beersma, B., & van Vianen, A. E. M. (2014). Does coaching work? A meta-analysis on the effects of coaching on individual level outcomes in an organizational context. *The Journal of Positive Psychology*, 9(1), 1–18. <https://doi.org/10.1080/17439760.2013.837499>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- 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). <https://arxiv.org/abs/2410.03017>