Performance Without Learning? AI Assistance Modes and the Production of Educational Performance: Evidence from a Randomized Field Experiment[1]

Sining Wang[a], Scott Shane[a], Yufan Dong[a,*], Heyu Xiong[a], Jie Ning[a], Leonardo Madureira[a], Richard Boyatzis[a]

[a] Weatherhead School of Management, Case Western Reserve University, 11119 Bellflower Road, Cleveland, OH 44106, USA

We study how the mode of generative AI assistance affects the production of observable educational performance using a randomized field experiment embedded in introductory accounting and microeconomics courses at a selective private university. Students were randomized at the student–topic level to one of four conditions: no AI, AI that provides solutions, AI that clarifies mistakes after independent work, and interactive AI tutoring. Each session paired a practice stage, in which AI access varied by treatment, with an immediately subsequent quiz completed without assistance by all students. We also measured each student's performance on each topic on later course exams. Solution-providing and interactive AI raise practice scores by 16 and 7 percent, respectively, yet none of the AI modes raises quiz scores, and the quiz–practice gap widens significantly under both performance-enhancing modes. Moreover, none of the AI modes changes exam scores. An ex post exploratory analysis shows that the interactive tutor's null effect on exam scores conceals opposing effects by type of knowledge: the tutor raises exam scores on procedural topics and lowers them on strategic topics. Solution-providing AI also lowers students' perceived learning and overall satisfaction by 11 percent, while interactive AI leaves perceptions unchanged. Solution provision shifts time from practice to the quiz, whereas interactive tutoring adds time to practice. The results indicate that AI assistance primarily changes how observable performance is produced rather than measured learning, with direct implications for how AI-assisted instruction should be evaluated.

Keywords: generative AI, field experiment, education production, performance measurement, time allocation, higher education

JEL codes: I21, I23, J24, O33

[1]The study protocol was approved by the IRB of Case Western Reserve University; all participants provided informed consent.