What's old is new: Tutorial-style presentations and improved learning

John C. Duchi Will Hartog Dennis L. Sun

October 16, 2026

Abstract

The ubiquity of Large Language Models poses a challenge to student learning in homework-based college courses, where access to easy homework answers devalues the intended productive struggle. To address this concern, we propose Oxbridge-style weekly tutorials, where students present homework problems without notes in a small group with an instructor or TA. We conducted a randomized experiment in an introductory probability class to evaluate the effectiveness of tutorials, finding a significant positive effect of tutorial enrollment on exam grades. We discuss the implementation, challenges, and opportunities in this style of course design.

1 Introduction

The rise of artificial intelligence, particularly Large Language Models (LLMs), has upended the role of homework in STEM courses, including probability and statistics. Homework has traditionally served several purposes, among them measuring learning, eliciting effort, and inducing productive struggle and persistence outside of class. Growing evidence suggests that LLMs pose a threat to all of these. First, it appears that LLMs erode the connection between learning and homework scores, as students can now achieve higher scores without learning more (Bastani et al. 2025; Strömberg et al. 2026). Second, LLMs reduce cognitive effort and study time, yielding lower-quality reasoning, downstream problem solving, and even brain activity (Stadler et al. 2024; Rismanchian et al. 2026; Kosmyna et al. 2025). Finally, preliminary work suggests LLMs reduce persistence and hinder unassisted performance (Liu et al. 2026). The traditional approach of assigning and grading homework thus no longer achieves its intended goals when powerful LLMs can substitute for individual effort.

Nevertheless, the educational psychology literature demonstrates that practice testing and distributed practice present two of the most effective learning strategies (Dunlosky et al. 2013). Homework provides an efficient way to implement both of these strategies, since it allows students to test themselves outside of class, spreading studying over time. It therefore seems premature to abandon homework in the face of LLMs, but rather to revise assessment so that it achieves its intended goals.

We explore one solution inspired by an old practice: tutorials at Oxford University (and Cambridge University, where they are known as "supervisions"). In tutorials, whose particulars vary from implementation to implementation, students meet regularly in small groups (of 1–6 students) with a "tutor" to discuss course material (Ashwin 2005, 2006). In

STEM fields, students typically complete a set of problems before each tutorial, and the tutor leads a discussion during the tutorial based on their answers.

In Section 2, we describe how we adapted the Oxbridge-style tutorials at an American research university. We measure the effectiveness of these tutorials for learning using a randomized experiment in an introductory probability course, which we discuss in Sections 3 and 4. To our knowledge, these are the first rigorous quantitative evaluations of the tutorial system.

2 Tutorials

Our goal is to retain homework as a method for distributed practice, while migrating assessment to a setting where AI assistance cannot stand in for a student's understanding. We left the weekly problem set essentially unchanged, except that the write-up no longer carries any weight in the grade. Instead, students are assessed on a weekly oral presentation in a small-group tutorial. Therefore, homework becomes preparation for an assessment rather than the assessment itself—essentially, an instance of learning by preparing to teach (Kobayashi 2019)—and a student who outsources the preparation pays the price a few days later in the tutorial, rather than several weeks later in the exam. We describe each component of the design in turn.

2.1 Homework

We assigned a problem set of six problems each week, which students were required to submit, as in a standard course. For students in the control (non-tutorial) group, we graded homework as usual; for students in the tutorial group, we recorded submissions but did not return written feedback. We continued to require submission for several reasons. First, a due date paces work, meaning it encourages distributed practice rather than cramming before exams, mimicking the best features of classical homework (Dunlosky et al. 2013). Second, submission gives a useful commitment device by providing immediate consequences for not turning problem sets in (Ryan and Hemmes 2005).

2.2 Tutorial Sessions

Students were assigned to tutorials of 4 students, each led by a teaching assistant (TA) or the professor, which met for 50 minutes once per week after the homework due date. A tutorial consists of 4 problems, one per student, so that every student presents exactly once. Importantly, these problems were *variants* of problems from that week's homework, which the students had not seen before. The following variants, from the units on counting and on expectation, are representative.

Example 1 (Counting and cards)**:** In this problem, we evaluate combinatorics using card games. We tell students to assume cards in the deck are completely randomized.

Homework problem. In Texas Hold'em, each player is dealt 2 cards, and players share 5 cards in the center of the table. A player has a flush if at least 5 of the 7 total cards share a suit. What is the probability you get a flush in one hand of Texas Hold'em?

Tutorial variant. In Bridge, four players play in two-person partnerships, where each player is dealt 13 of the 52 cards in the deck. A partnership is long in a suit if it has at least 10 cards in that particular suit between its two players. What is the probability your partnership is long in a suit in one hand of Bridge?

Both problems require the same decomposition into suits and the same summation for the union of (e.g.) 5, 6, or 7 cards of a suit, but no verbatim calculation of the Hold'em solution answers the Bridge question. $\diamond$

Example 2 (Expectations of functions)**:** We asked students to compute expectations $\mathbb{E}[g(X)]$ for different functions g. In each variant, X was a random variable taking values in $\{2^k\}_{k=1}^{\infty}$, with $\mathbb{P}(X = 2^k) = 2^{-k}$ for $k = 1, 2, \ldots$.

Homework problem. Compute $\mathbb{E}[\log_2 X]$.

Tutorial variant. Compute $\mathbb{E}[\sqrt{X}]$.

Typical misconceptions swap $\mathbb{E}[g(X)]$ for $g(\mathbb{E}[X])$, leading to confusion when $\mathbb{E}[X] = +\infty$; a student who has internalized the distinction handles either expectation easily. $\diamond$

A student who worked and understood the original problem should find the variant straightforward; a student who only memorized the solution may transfer little.

To encourage students to study all problems, tutors did not reveal which problem a student would present until it was that student's turn, and students worked at the board without notes. Those not presenting were asked to stay off devices and to follow the presenter's argument; when a presenter was stuck, we encouraged the others to offer suggestions. In our experience, this inclusion of other students helps to mitigate the stress of assessment, making clearer that we view the tutorials as something closer to a working session than examination.

2.3 Grading

Perhaps the biggest difference between our implementation of tutorials and those at Oxbridge is that our tutorials were graded. We did this to provide incentives similar to those of traditional graded homework. However, the grade was based on perceived effort, engagement, and participation, rather than mastery (which would be measured on in-class exams). The evaluation centered on whether they had clearly attempted the homework, whether they engaged with the problem at the board and with the instructor's questions, and whether they attended to and helped their peers. We hope that this evaluation attaches the grade to working the homework without outsourcing it, measuring the learning process rather than the product.

Instructors recorded a quantized score in $\{0, 1, 2\}$ for each tutorial, which replaced the homework component of the course grade. A coarse three-point scale (i) requires only a few seconds to assign, (ii) admits little argument over boundary cases, and (iii) signals that the tutorial is not an exam. A student who prepares but gets stuck on the variant problem receives the same score as a student who sails through it, encouraging productive struggle.

2.4 Instructions to Tutors

We provided tutorial leaders with instructions and materials to standardize student experience within the tutorial sessions. The teaching team provided a roadmap for each potential problem in the weekly problem set consisting of details to ask the students, including follow-up questions in case time runs long. Tutorial leaders filled any remaining time with discussion of the other problems if students finished early. We also gave tutorial leaders the following general principles:

- Students should be attentive to their peers when not presenting; they should not be reviewing notes or trying to memorize homework. As such, the tutorial leader should not let the students know their assigned problem until it is their turn to present.
- The tutorial is not meant to be a quiz. It is for building concepts and reinforcing learning, and tutorial grades are based not on correctness but preparedness and effort.

3 Experimental Design

To evaluate the tutorials of Section 2, we implemented a randomized controlled experiment, with blocking based on pretest performance. We describe the setting, the treatment and control conditions, and our stratification below.

3.1 Setting

Each subject in this study was a Stanford student enrolled in Stats 117, Introduction to Probability, in Spring Quarter 2026. The course covers concepts in a standard probability course, with major units on foundations of probability, counting, discrete random variables, expectation and variance, conditioning, and continuous random variables. The 10-week course consisted of three 50-minute lectures per week, with one problem set assigned as homework each week. Students in both conditions attended the same lectures and sat the same in-class midterm and final examinations, which supply the outcome measures for the analysis of Section 4.

3.2 Treatment group

Students assigned to the treatment group participated in the weekly tutorials of Section 2. As described there, they still handed in their written homework, but it was not graded; they were evaluated exclusively on their tutorial performance, which took the place of the homework component of their course grade.

3.3 Control group

Students in the control group participated in an essentially standard college-level course, at least as offered in the United States. They turned in written homework for grading, and the course staff advised them against using AI assistance, suggesting that it was likely to hinder their performance on the midterm and final exams. Weekly 50-minute sections were offered by TAs, and students were asked to attend one section each week. These sections

consisted of live-solving of problems involving either concepts or examples from the weeks' lectures, led by TAs. Students who attended section each week spent exactly the same time as students in tutorials. However, we did not enforce attendance in these sections, so students in the control group who chose not to attend had less time-on-task than students in the treatment group.

3.4 Blocking and control

To reduce variation due to differences in student background, we blocked based on performance on a concept inventory pre-test, which was issued on the first day of class. This inventory consisted of 12 total questions that evaluated students' conceptual and intuitive understanding of probability, related to (i) independence, (ii) sequential sampling (geometric random variables), (iii) sampling with and without replacement, (iv) expected values, (v) and conditioning.

We then blocked students based on their pretest score and randomized students with the same score so that approximately half were in treatment and half were in control. So, for example, among the 5 students scoring 12/12 on the pretest, 3 were in tutorials and 2 were in the control group, and among the 10 students scoring 9/12 on the pretest, 5 were in the tutorials and 5 in the control.

4 Results and analysis

We present summary results in this section for students who consented to have their data included in the study. Of the 80 students who enrolled in the class, all $n = 66$ who responded to the form consented to the use of their data. The remaining students who did not respond may result in some selection bias; while we cannot present the full results, they remain consistent with the conclusions here. We present two analyses: a standard fixed-effects analysis of a randomized block design and a fully nonparametric randomization test. All inferences (p-values, confidence intervals) are one-sided.

4.1 Fixed-effects analysis of a randomized block design

Let $W_i \in \{0, 1\}$, $X_i \in \{-1, 0, 1, \ldots, 12\}$, and Y_i denote the treatment assignment, pretest score, and outcome (e.g., midterm or final score), respectively, of student i, where we code $X_i = -1$ to indicate that the student did not complete the pretest. We fit the standard linear model

$$Y_i = \tau 1\{W_i = 1\} + \sum_{k=-1}^{12} \beta_k 1\{X_i = k\} \tag{1}$$

to the exam scores. In Model (1), the average treatment effect τ constitutes the parameter of interest, and our randomization guarantees that least squares is consistent for this quantity (Imbens and Rubin 2015). Table 1 summarizes the results, showing estimated treatment effects of 14.9 percentage points for the midterm and 11.4 percentage points for the final ($p < 0.002$), with 95% lower confidence bounds for the effect size τ (in percentage points) of 6.84 percentage points the midterm and 5.30 percentage points on the final.

Outcome	n	$\widehat{\tau}$	std. error	t-score	p-value	95% lower τ
Midterm	66	14.9	4.84	3.08	0.002	6.84
Final	64	11.4	3.66	3.12	0.001	5.30

Table 1. Results of the fixed-effects analysis by fitting Model (1), measuring effect sizes in percentage points. Two students dropped the course after the midterm examination and did not take the final.

4.2 Randomization test

Under the null hypothesis that the treatment has no effect, the distribution of any test statistic is invariant to permutations of the treatment assignments. Thus, the permutation null—permuting treatment assignments and recalculating the test statistic—provides a reference null distribution for any test. (We permute treatment assignments separately for each pretest score to mirror the randomized block design of the study.) One obtains a p-value by locating the observed difference within this reference null distribution.

So long as the test statistic is monotone in outcome differences, the randomization test remains valid for testing the one-sided null hypothesis

$$H_0 : Y_i(1) \leq Y_i(0) \quad \text{for each } i, \tag{2}$$

where $Y_i(t)$ denotes the *potential outcome* (Imbens and Rubin 2015) of subject i under treatment t (Caughey et al. 2023). Under the null hypothesis (2), participation in tutorials has no positive effect on any individual's midterm or final performance, and this p-value is exact, without any assumptions on the distribution of the potential outcomes. Following Caughey et al. (2023), we also identify the smallest τ for which the hypothesis

$$H_\tau : Y_i(1) \leq Y_i(0) + \tau \quad \text{for each } i \tag{3}$$

cannot be rejected at level $\alpha = .05$. This provides the natural analog of the "95% lower confidence bound" in the fixed-effects analysis.

Table 2 summarizes results, yielding estimated treatment effect (difference in means between the two groups) of 14.3 percentage points for the midterm and 9.5 percentage points for the final ($p < 0.002$). The "95% lower confidence bound", in the sense of (3), is 6.74 precentage points and 5.19 percentage points.

Outcome	n	Tutorial mean	Control mean	$\widehat{\tau}$ (pp)	p-value	95% lower τ
Midterm	66	72.7	58.4	14.3	0.002	6.78
Final	64	59.5	50.0	9.5	0.002	5.19

Table 2: Results of the randomization test in percentage points.

5 Conclusions

The significant positive effects we observe in our randomized experiment suggest that tutorials are a potentially powerful educational invention. However, we cannot identify the precise causal mechanism in this study. We hypothesize that at least two factors coincide

here: in addition to the personalized instruction that tutorials afford, AI may also be depressing learning from traditional homework. If this hypothesis is correct, then tutorials are not just a useful educational intervention but one that is uniquely suited to mitigating the detrimental effects of AI on learning.

A natural concern that tutorials raise is their cost: how is it possible to achieve ratios of 4 students per TA? In order for tutorials to be feasible, a university needs to have ratios of under 4 students per TA-hour. The ratio needs to be under 4 to allow for preparation and administrative tasks, but tutorials also reduce or eliminate the need for other traditional TA duties, such as grading homework and some office hours (since students have the opportunity to ask questions in tutorial). At Stanford, a private research university, the ratio is 1.5–2.0, but this ratio varies widely across institutions. At large public research universities, a brief survey suggests the ratio ranges from roughly 3.0, at the University of Michigan (Gala-Garza 2024), UT Austin (McMillan 2026, Tab. 2), and UCLA (O'Leary et al. 2021, "Secondary Sections"), to 4.0 at UCSD (Workgroup on the Future of Graduate Education 2026, Sec. 4.5) and Oregon State University, College of Earth, Ocean, and Atmospheric Sciences (2026).

As a practical matter, even if a TA with a 20-hour-per-week appointment could technically hold 15 tutorials, that may be exhausting. Our TAs had shorter, 10-hour-per week appointments, meaning that each TA only needed to hold 4 tutorials each week (i.e., 2 days with 2 tutorials each). We have also had success hiring senior undergraduate tutorial leaders, who excelled in a previous offering of the class and were highly recommended by their tutorial leader.

Many questions remain. Most saliently, our study focuses on a single course on introductory probability at Stanford, with a single instructor. Because we encouraged, but did not enforce, the control group to attend sections, time on task or additional time with instructors may influence the observed improvements. Moreover, our study does not disentangle more precise causative mechanisms: the tutorials include unseen variant problems, require explaining solutions at the board, allow small-group participation, and use effort-based grading. Larger classes may necessitate changes to the format, such as asynchronously scheduled one-on-one or remote sessions, and it is unclear whether the observed benefits would survive these changes. We are currently investigating these questions in a larger study across several courses in several different disciplines.

Even with these obvious caveats, our results suggest that tutorials provide a promising response to new technology. For centuries, tutors at Oxbridge have asked students to explain their work, and when a written solution may no longer reflect what a student understands, this approach may provide precisely those incentives that keep students learning. While an LLM can type a student's homework, it cannot put chalk to the chalkboard for them.

6 Disclosure statement

The authors have no conflicts of interest.

7 Data Availability Statement

Because of the sensitivity of personally identifiable data in the analyses, we do not make data available. The supplement includes the analysis pipeline as well as hashes of the raw data analyzed to allow verification by approved individuals. This study was approved by the Stanford Institutional Review Board, IRB #83954.

A Analysis details

A.1 Analysis of the randomization tests

For observed scores Y_i^{obs} and a vector $\boldsymbol{Z} = (Z_i)_{i=1}^n \in \{0,1\}^n$ of treatment indicators and associated counts $n_z = \sum_{i=1}^n 1\{Z_i = z\}$, $z = 0, 1$, we consider the sample mean difference

$$T_{\boldsymbol{Z}} \coloneqq \frac{1}{n_1} \sum_{i:Z_i=1} Y_i^{\text{obs}} - \frac{1}{n_0} \sum_{i:Z_i=0} Y_i^{\text{obs}},$$

so that for the actual assigments $\boldsymbol{Z}^{\text{obs}}$ we have $T_{\boldsymbol{Z}^{\text{obs}}} = \frac{1}{n_1} \sum_{i:Z_i^{\text{obs}}=1} Y_i(1) - \frac{1}{n_0} \sum_{i:Z_i^{\text{obs}}=0} Y_i(0)$. To evaluate the null (2), we consider all randomizations $\boldsymbol{Z}^* \in \{0,1\}$ respecting the pretest scores and initial counts, so that $\boldsymbol{Z}^*$ are sampled uniformly over the collection of treatment assignments with number of individuals in each pre-test strata equal to that in $\boldsymbol{Z}^{\text{obs}}$. As Caughey et al. (2023) show, the randomization probability $\mathbb{P}(T_{\boldsymbol{Z}^{\text{obs}}} \leq T_{\boldsymbol{Z}^*} \mid Y^{\text{obs}})$ is a valid p-value for the weaker one-sided null that $Y_i(1) \leq Y_i(0)$ for each i. We also identify the smallest value τ at which we can reject the null

We estimate the value $\widehat{\tau}$ in the null (3) at level α as follows. Imputing at the boundary $Y_i(1) = Y_i(0) + \tau$ yields $Y_i(0) = Y_i^{\text{obs}} + \tau Z_i^{\text{obs}}$ and $Y_i(1) = Y_i^{\text{obs}} - \tau(1 - Z_i^{\text{obs}})$, giving the adjusted statistic

$$T_{\boldsymbol{Z},\tau} \coloneqq T_{\boldsymbol{Z}} + \tau \left(\frac{1}{n_1} \langle \boldsymbol{Z}^{\text{obs}}, \boldsymbol{Z} \rangle - \frac{1}{n_0} \langle \boldsymbol{Z}^{\text{obs}}, \mathbf{1} - \boldsymbol{Z} \rangle \right) = T_{\boldsymbol{Z}} - \tau \left(1 + \left(\frac{1}{n_1} - \frac{1}{n_0} \right) \langle \boldsymbol{Z}^{\text{obs}}, \boldsymbol{Z} \rangle \right).$$

Caughey et al. (2023) show that $\mathbb{P}(T_{\boldsymbol{Z}^{\text{obs}}} \leq T_{\boldsymbol{Z}^*,\tau} \mid Y^{\text{obs}})$ is a valid p-value for the null (3). To compute p-values, we use a Monte-Carlo simulation of $N = 10^6$ resamples of $\boldsymbol{Z}$, computing the one-sided p-value $p = \frac{1}{N} \sum_{i=1}^N 1\left\{ T_{\boldsymbol{Z}^{\text{obs}}} \leq T_{\boldsymbol{Z}_i^*} \right\}$. Because $\tau \mapsto T_{\boldsymbol{Z},\tau}$ is monotonically decreasing, we can efficiently solve for $\widehat{\tau}$, the largest τ at which we can reject the null (3) at level α by solving

$$\widehat{\tau} \coloneqq \sup \left\{ \tau \mid \alpha > \frac{1}{N} \sum_{i=1}^N 1\left\{ T_{\boldsymbol{Z}^{\text{obs}}} \leq T_{\boldsymbol{Z}_i^*,\tau} \right\} \right\}.$$

SUPPLEMENTARY MATERIAL

We provide a full analysis pipeline for reproduction of our results, assuming access to raw data files, on the public github for this paper. To allow verification of the raw data being used, we include the following SHA256 hashes of our data files, also described in the `PIPELINE.md` file there:

- `raw-sp26-gradescope-grades.csv`
  ac077d0483a4b4ef36be07fc6073e65224df938c658ae7bf2a3226da09d1e727
- `raw-tutorial-assignments.csv`
  1c7d7e0cadca32c7c90763e19c0bcdb48324747ac8760828a6ecb0dd0d83f7d5
- `raw-consent.csv`
  1d89697dcd90b76c3244997991e135c0b31a9856d09f67df450003931f76693c
- `raw-pretest-scores.csv`
  a051dea5c5aec15b36c204285a2b3a8b71219e74e7cc85c1f4e697eb9c40686b

References

P. Ashwin. Variation in students' experiences of the "Oxford tutorial". *Higher Education*, 50(4):631–644, 2005.

P. Ashwin. Variation in academics' accounts of tutorials. *Studies in Higher Education*, 31 (6):651–665, 2006.

H. Bastani, O. Bastani, A. Sungu, H. Ge, Özge Kabakcı, and R. Mariman. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26):e2422633122, 2025.

D. Caughey, A. Dafoe, X. Li, and L. Miratrix. Randomisation inference beyond the sharp null: bounded null hypotheses and quantiles of individual treatment effects. *Journal of the Royal Statistical Society, Series B*, 85:1471–1491, 2023.

J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest*, 14(1):4–58, 2013.

T. Gala-Garza. LSA reduces GSI allocations by 8% for fall 2024. *The Michigan Daily*, 2024. URL https://www.michigandaily.com/news/academics/lsa-reduces-gsi-allocations-by-8-for-fall-2024/.

G. Imbens and D. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

K. Kobayashi. Learning by preparing-to-teach and teaching: A meta-analysis. *Japanese Psychological Research*, 61(3):192–203, 2019.

N. Kosmyna, E. Hauptmann, Y. T. Yuan, J. Situ, X.-H. Liao, A. V. Beresnitzky, I. Braunstein, and P. Maes. Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv:2506.08872 [cs.AI]*, 2025.

X. Li and P. Ding. General forms of finite population central limit theorems with applicationsto causal inference. *Journal of the American Statistical Association*, 112(520): 1759–1769, 2017.

G. Liu, B. Christian, T. Dumbalska, M. A. Bakker, and R. Dubey. AI assistance reduces persistence and hurts independent performance. *arXiv:2604.04721 [cs.AI]*, 2026.

H. Liu and Y. Yang. Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika*, 107(4):935–948, 2020.

A. McMillan. Soft money instructional funds allocation guidelines. The University of Texas at Austin, College of Liberal Arts, 2026. URL https://utexas.atlassian.net/wiki/spaces/larc/pages/75241069/Soft+Money+Instructional+Funds+Allocation+Guidelines. Accessed October 14, 2026.

E. S. O'Leary, H. W. Sayson, C. Shapiro, A. Garfinkel, W. J. Conley, M. Levis-Fitzgerald, M. K. Eagan, and B. Van Valkenburgh. Reimagining the introductory math curriculum for life sciences students. *CBE—Life Sciences Education*, 20(4):ar62, 2021. doi: 10.1187/cbe.20-11-0252. URL https://doi.org/10.1187/cbe.20-11-0252.

Oregon State University, College of Earth, Ocean, and Atmospheric Sciences. CEOAS GTA allocation guidelines, 2026. URL https://ceoas.oregonstate.edu/gta-allocation-guidelines. Accessed October 14, 2026.

S. Rismanchian, H. Uzun, J. Matayoshi, E. Cosyn, and E. Kurd-Misto. Faster completion, less learning: Generative AI reduced study time on math problems and the knowledge they build. *arXiv:2605.21629 [cs.CY]*, 2026.

C. S. Ryan and N. S. Hemmes. Effects of the contingency for homework submission on homework submission and quiz performance in a college course. *Journal of Applied Behavior Analysis*, 38(1):79–88, 2005. doi: https://doi.org/10.1901/jaba.2005.123-03. URL https://onlinelibrary.wiley.com/doi/abs/10.1901/jaba.2005.123-03.

M. Stadler, M. Bannert, and M. Sailer. Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 160 (108386):1–7, 2024.

D. Strömberg, V. Lei, and Y. Wu. The generative AI learning penalty: Evidence from Chinese secondary education. *Centre for Economic Policy Research*, (DP21577), 2026. Discussion Paper.

Workgroup on the Future of Graduate Education. Final report. Technical report, University of California, San Diego, June 2026. URL https://senate.ucsd.edu/media/768814/sawg-on-future-of-graduate-education-review-docs.pdf.