NET WORK LISER

Discussion Paper Series

IZA DP No. 18513

April 2026

Guidance Over Adoption: Experimental Evidence on AI-Assisted Learning

Sebastian Gallegos
UAI Business School, J-PAL, IZA@LISER
and HCEO, University of Chicago

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18513 April 2026

Guidance Over Adoption: Experimental Evidence on AI-Assisted Learning*

Abstract

This paper estimates the causal effect of a large language model–based study assistant on student behavior and learning outcomes in a natural field setting with real academic stakes. I design and deploy a course-specific AI assistant (GPT-UAI) for undergraduate econometrics and evaluate it through two randomized interventions implemented across seven coordinated course sections at a selective university in Chile. The first intervention targets the extensive margin of use, encouraging GPT-UAI adoption prior to the midterm exam. The encouragement raises the GPT's awareness and reported usage, but does not change its perceived value and does not improve midterm performance. The second intervention targets use at the intensive margin, providing guidance on learning-oriented usage for the final exam. Guidance shifts interactions with GPT-UAI toward tutor-style engagement, increases perceived usefulness by 0.38 standard deviations, improves final-exam performance by 0.21 standard deviations, and raises the probability of earning a passing exam grade by 12 percentage points. The findings suggest that learning gains arise less from adoption than from guiding how students use course-specific AI assistants.

JEL classification

I23, C93, O33, D83

Keywords

generative AI, large language models, higher education, field experiments, randomized controlled trials, student learning, human capital, AI-assisted learning, tutoring, technology in education

Corresponding author

Sebastian Gallegos
sebastian.gallegos@uai.cl

* I gratefully acknowledge financial support from the Teaching Innovation Research Fund 2025 (Fondo de Investigación en Docencia) at Universidad Adolfo Ibáñez (UAI). This study was pre-registered at the American Economic Association Registry for Randomized Controlled Trials (AEARCTR-0016770; https://doi.org/10.1257/rct.16770-1.0) and at the Center for Open Science (OSF Registry; https://doi.org/10.17605/OSF.IO/ZNMRJ). Sofia Inostroza provided research assistance. I thank Alejandro Montecinos and Diego Avanzini for facilitating access to the encrypted administrative data. Diego Avanzini also helped with the implementation of the intervention. I am also grateful to my colleagues Nieves Valdés, Bernardo Lara, Nicolás Libuy, and Felipe Morales for helpful comments and for their willingness to participate in the implementation of the study with their econometrics sections. All remaining errors are my own.

1 Introduction

Generative artificial intelligence has entered university education faster than any previous instructional technology. By late 2025, AI-based study tools had become a standard component of higher education. Amoah et al. (2025) report that by mid-2024, already approximately 70% of higher-education students worldwide were already using ChatGPT, the large language model (LLM) developed by OpenAI. Despite this rapid diffusion, rigorous causal evidence of the adoption and use of LLMs in real-world higher education settings is still scarce. Universities are making policy decisions about whether to restrict, regulate, or integrate these tools (Jin et al., 2025), yet empirical evidence to inform these decisions remains limited.

This paper provides causal evidence on the effects of an LLM-based study assistant on student behavior and learning in a natural field setting with real academic stakes. I design and deploy a course-specific assistant (GPT-UAI) trained on materials from a standard undergraduate econometrics curriculum. I then assess its adoption and use through two independent randomized controlled trials (RCTs) implemented during the second semester of 2025 across seven coordinated course sections at a selective university in Chile, enrolling about 300 students.[1]

The research design has three main features. First, it shifts the extensive and intensive margins of AI use through two interventions, one that encourages adoption prior to the midterm exam and another that provides guidance on learning-oriented use before the final exam. Second, it combines experimental variation with administrative and survey data. Administrative records provide objective measures of academic performance, while survey data capture adoption, usage, perceived value, and study behavior. Third, the design allows for and measures peer spillovers, reflecting how information diffuses across students in real-world settings.

The encouragement intervention increases awareness (knowing about GPT-UAI) and reported usage by 21 and 11 percentage points, respectively. However, it does not change its perceived value and does not improve midterm performance. Effects on grades and scores are small and statistically indistinguishable from zero.

In contrast, the guidance intervention improves final exam performance by 0.21 standard devia-

[1]This study was pre-registered at the American Economic Association Registry for Randomized Controlled Trials (AEARCTR-0016770; https://doi.org/10.1257/rct.16770-1.0) and at the Center for Open Science (OSF Registry; https://doi.org/10.17605/OSF.IO/ZNMRJ).

tions and raises the probability of earning a passing exam grade by 12 percentage points. Effects are concentrated in open-response questions requiring structured reasoning, with no detectable impact on multiple-choice questions. The learning gains appear to operate through changes in how students use and value the GPT. The guidance intervention increases perceived usefulness by 0.38 standard deviations, shifts interaction toward more tutor-style engagement by 10 percentage points, and raises the relative importance of the tool within students' study portfolio by 24%. Taken together, the results from both interventions suggest that improvements in academic performance depend less on adoption of the course-specific GPT alone and more on guidance and usage quality.

These findings are specific to this setting, but reflect an environment that is increasingly common in higher education elsewhere. Access to generative AI is now widespread, and higher education students routinely use tools such as ChatGPT, Gemini, Claude, or Copilot. The relevant question is therefore not whether students use AI or not, but how these tools are integrated into the learning process.

This paper is among the first to provide causal evidence on LLM-based study assistants in higher education, contributing to different strands of the literature. First, it contributes to the emerging literature on generative AI in education. Relevant recent work in higher education includes Kestin et al. (2025) and Fischer et al. (2025), who study AI tutoring in stylized environment settings, and Contractor and Reyes (2025), who document rapid adoption and heterogeneous usage among college students. In primary and secondary education, randomized evaluations in Turkey, Nigeria, and India show that AI-based tools can improve learning outcomes under guided interaction (De Simone et al., 2025; Agrawal et al., 2026). This paper complements this evidence with a design from a natural field setting with real stakes in higher education, and by distinguishing between adoption and usage as separate margins of AI use.

Second, the paper contributes to the literature on technology and education tutoring. Prior evidence on technology-based inputs is sobering, with computer access programs generally yielding small or null effects on learning outcomes (Malamud and Pop-Eleches, 2011; Cristia et al., 2017; de Barros and Ganimian, 2024; Cueto et al., 2025). More recent work shows that interventions that shape how technology is used can be more effective (Muralidharan et al., 2019; Rodriguez-Segura, 2021). This distinction parallels the long-standing evidence on tutoring, where one-on-one instruction generates large learning gains (Bloom, 1984; Nickow et al., 2020), motivating the search

for scalable alternatives. Generative AI tools are a natural candidate, as they provide continuous, low-cost access to individualized support.

Overall, this paper contributes with new evidence to a literature that is set to increase rapidly in the following years. The arrival of LLMs alters the process of human capital accumulation by providing students with near-costless, always-available study assistance. At the same time, these AI tools can either support or substitute for the underlying cognitive processes that generate learning. As generative AI becomes further integrated into higher education, an important direction for future work is to understand how guidance interacts with different learning environments and modes of use.

2 The Setting and the GPT-UAI tool

The AI Assistant. I built the GPT EconometriaUAI (GPT-UAI onwards) within the OpenIA ChatGPT platform to provide students with an assistant aligned with the content and difficulty of the econometrics course. The model was trained on instructional materials from ten previous semesters, including lecture notes, problem sets, exams, and solutions.[2]

The GPT-UAI interface displays a standardized message indicating that the assistant was built using materials from the undergraduate econometrics course at the UAI Business School. Figure A.1 shows a screenshot of the interface. The welcome message explains that the tool is an assistant for studying econometrics trained with course materials from the course instructor, and that responses should be considered indicative rather than definitive. It also encourages students to consult teaching assistants or the course instructor for clarification when needed.

The GPT-UAI was accessible through the standard (free) ChatGPT interface using a shared link, with no paywalls or institutional restrictions. It was introduced as a complement to lectures and teaching assistant sessions, and did not track individual usage, reflecting typical conditions under which students use AI tools.

Course and Institutional Context. The course follows a standard undergraduate econometrics curriculum, covering core topics such as ordinary least squares, inference, and introductory causal methods.[3] During the second semester of 2025, it was offered in seven coordinated sections of 40–50 students each, all following a common syllabus and using shared assessment materials.

Students are enrolled in either Economics or Business Administration within a "3+2" degree structure that combines a three-year bachelor's phase with a two-year master's specialization. During the bachelor phase, students complete core coursework and enter the master's stage upon meeting academic requirements. Econometrics is a required course typically taken in the third year and is considered one of the most demanding in the curriculum. Passing the course is necessary to enter the master's stage, making the midterm and especially the final exam high-stakes assessments.

[2]The assistant was used in a pilot phase during the previous academic year without formal evaluation.
[3]The full syllabus is included in Appendix B.

3 Experimental Design

This section describes the experimental procedures and the two randomized interventions implemented during the second semester of 2025 (August–December). The first intervention took place at the end of September, before the midterm exam. The second intervention occurred at the end of November, before the final exam.

3.1 RCT 1: Encouragement (Extensive Margin)

Intervention. The first intervention targeted adoption of the course-specific GPT, encouraging its use at the extensive margin. It was delivered to students in the treatment group through the official course platform (*Webcursos*).[4] Treatment students received three official encouragement messages from the Econometrics Coordination team. The first message introduced the GPT-UAI and provided the access link. The second presented a concrete example of how to use it for exam preparation. The third served as a reminder and included a link to a short instructional video. These messages were delivered in the days preceding the midterm exam.[5]

Students assigned to the control group did not receive encouragement emails but could access the GPT through *Webcursos*. This reflects university rules, which did not allow restricting access to the tool. Instructors did not promote the GPT-UAI beyond making it available.

Randomization. The randomization sample consisted of all students enrolled in undergraduate econometrics at the beginning of the semester. I randomized them at the individual level within strata using baseline covariates available in the administrative data, following recommended practice in the experimental literature (Duflo et al., 2007; Athey and Imbens, 2017; Bruhn and McKenzie, 2009). I formed strata based on section, gender, and whether the student attended a private or non-private high school. [6] I assigned one-third of students to treatment and two-thirds to control. This allocation was chosen to limit contamination of the control group, although the design explicitly allows for spillovers, as we explain below and in the Empirical Strategy section.

[4] *Webcursos* is the official learning management system of UAI (webc.uai.cl), analogous to Canvas or Blackboard. It hosts course materials, announcements, and academic records.

[5] Appendix C reproduces the exact timing and full content of the messages.

[6] I limited stratification to this set of covariates because other pre-treatment variables in the administrative data exhibited little variation (e.g., age, year of entry) or would have required arbitrary discretization (e.g., degree progression). Expanding the set of stratification variables would have produced many sparsely populated strata without clear gains in balance. All empirical specifications control for the stratum variables.

3.2 RCT 2: Guidance (Intensive Margin)

Intervention. The second intervention targeted how students used the GPT-UAI rather than whether they used it. Students assigned to treatment received three emails prior to the final exam encouraging learning-oriented use of the tool. The messages promoted tutor-style interaction with step-by-step reasoning, provided structured prompts to generate practice questions and feedback, and encouraged verification of answers.[7]

Control students did not receive guidance emails but retained full access to the GPT, as in RCT 1. To measure compliance and spillovers, the final exam survey asked whether students received the emails and whether they shared the advice.

Randomization. I implemented a second randomization prior to the final exam using the same sample of originally enrolled students. This randomization was independent of RCT 1 and followed the same stratification scheme. Within each stratum, approximately half of the students were assigned to treatment and half to control.

[7] Appendix D reproduces the full emails and their timing.

4 Data and Measurement

Administrative Data and Learning Outcomes. Administrative records provide baseline covariates and learning outcomes. Baseline variables include the stratification variables (gender, section, and high school type), as well as date of birth, year of entry, specialization (Economics or Business Administration), and degree progression (0-100 scale).

I measure learning outcomes using midterm and final exam performance. I use standardized grades (normalized to mean zero in the control group), raw grades on the 1–7 scale, and exam scores from 0 to 100. The Economics exam is fully open response, while the Business Administration exam includes both multiple-choice and open-response components, for which I observe separate subscores.

Survey Measures and Adoption Variables. I collected survey data at two points in the semester, embedded in the midterm and final exams.[8] These surveys measure adoption, usage, perceptions, and spillovers.

The midterm survey focuses on the encouragement intervention. It asks whether students know about GPT-UAI, how they learned about it, whether they used it, and whether they recall receiving the encouragement emails. From these responses, I construct indicators for awareness (knowing about GPT-UAI), usage (using the GPT at least once), and usage intensity, measured as the self-reported number of times the GPT was used. The survey also measures perceived usefulness on a 0–10 scale, from which I construct indicators for high usefulness (7–10) and very high usefulness (10).

The final exam survey focuses on the guidance intervention. It asks whether students received the guidance emails, whether they shared the guidance prompts with peers, and collects information on GPT-UAI usage and perceived usefulness, as in the midterm survey. I construct an indicator for guidance exposure (equal to one for students reporting receipt of guidance through emails or peers), and measures of GPT-UAI use and intensity analogous to those in the midterm survey. The survey also asks about the use of other AI tools (ChatGPT, Claude, Gemini, Copilot, etc.) and about the mode of GPT-UAI use on a 0–5 scale, from which I construct indicators for tutor-style use (4–5) and full tutor mode (5). Finally, students rate the usefulness (on a 1–5 scale) of several course

[8]The surveys were physically printed and stapled to the corresponding exam. The questions were not graded and participation was voluntary. At the start of each exam, instructors in each section asked students to spend approximately five minutes completing the survey before the exam formally began.

resources during the semester, including GPT-UAI, teaching assistant sessions, in-person lectures, lab sessions, problem set guides, and readings.

Sample and Balance. The analytic samples consist of 303 students for RCT 1 and 289 for RCT 2, corresponding to approximately 80% and 75% of all students enrolled in undergraduate econometrics at the beginning of the semester. Participation in each analytic sample is balanced by treatment status in both experiments, and baseline individual covariates do not predict inclusion in either RCT, although participation rates vary slightly across course sections.[9]

Table 1 reports summary statistics for students participating in each RCT. The mean age is 21.8 years and 31 percent are women. The average entry year is close to 2023, implying that by the time of the intervention (second semester of 2025) students are roughly halfway through their third year. Consistent with this, the mean degree progression rate is 67 percent. About 75 percent attended private high schools, and close to 70 percent are in Business Administration.

Treatment and control groups are balanced across baseline administrative characteristics in both RCTs. Table 1 shows that differences by treatment status in both experiments are small with no statistically significant differences and joint tests failing to reject equality of covariates.

[9] See Table F.5 and Table F.6.

Table 1: Baseline Characteristics and Balance

	RCT 1				RCT 2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	All Mean	Control Mean	Mean Diff.	p-value	All Mean	Control Mean	Mean Diff.	p-value
Female	0.310	0.327	-0.050	0.374	0.308	0.309	-0.002	0.977
Private High	0.759	0.748	0.035	0.500	0.737	0.698	0.081	0.120
Degree Progression	66.973	66.470	1.509	0.430	67.318	66.624	1.435	0.433
Major Business	0.700	0.693	0.020	0.722	0.671	0.678	-0.014	0.807
Entrance Cohort	2022.941	2022.985	-0.134	0.375	2022.927	2022.879	0.099	0.446
Exact Age	21.800	21.689	0.333	0.081	21.819	21.833	-0.027	0.881
Birth Year	2003.389	2003.490	-0.302	0.122	2003.377	2003.349	0.058	0.754
Birth Month	6.739	6.851	-0.337	0.418	6.654	6.812	-0.326	0.412
Birth Day	15.581	15.812	-0.693	0.507	15.588	16.255	-1.376	0.169
Section 1	0.152	0.144	0.025	0.581	0.152	0.154	-0.004	0.918
Section 2	0.129	0.129	0.000	1.000	0.173	0.181	-0.017	0.705
Section 3	0.165	0.153	0.035	0.457	0.138	0.154	-0.033	0.418
Section 4	0.119	0.124	-0.015	0.702	0.104	0.087	0.034	0.345
Section 5	0.135	0.144	-0.025	0.543	0.104	0.101	0.006	0.858
Section 6	0.168	0.173	-0.015	0.743	0.194	0.195	-0.002	0.970
Section 7	0.132	0.134	-0.005	0.904	0.135	0.128	0.015	0.704
	All	Control	Treat.	Joint test	All	Control	Treat.	Joint test
Observations and p-value of joint test	303	202	101	0.814	289	149	140	0.640

Notes: Table 1 reports baseline characteristics and balance for the two randomized interventions. For RCT 1, columns (1)–(4) report the overall mean, the control group mean, the treatment–control mean difference (labeled "Mean Diff." and estimated from a regression of the characteristic on the treatment indicator), and the corresponding p-value. Columns (5)–(8) report the same statistics for RCT 2. The bottom panel reports the number of observations for each RCT (overall, control, and treatment) and the p-value from a joint test of the baseline variables obtained from a regression of all characteristics on the corresponding treatment indicator.

5 Empirical Strategy

Reduced-form Effects. For each RCT, I estimate reduced-form effects on adoption and usage measures derived from the survey data, and on learning outcomes as described in the Data and Measurement section. The corresponding main estimating equation is:

$$Y_i^j = \alpha^j + \beta^j R_i^j + \gamma^j X_i + \lambda_s + \varepsilon_i^j \tag{1}$$

where $j = 1, 2$ indexes the randomized interventions. Y_i^j denotes an outcome for student i measured for RCT j, R_i^j indicates assignment to treatment in RCT j, X_i includes predetermined covariates, λ_s represents strata fixed effects, and ε_i^j is an idiosyncratic error term for each RCT. The coefficient β^j captures the intention-to-treat effect randomized intervention j.

Spillovers and Endogeneity. Both interventions provide information that could diffuse from treated to control students, making exposure to the relevant mechanisms endogenous. Rather than attempting to eliminate such spillovers, the design allows and measures them. I instrument each exposure with the corresponding randomized assignment.

For RCT 1, D^1 equals one if student i reports awareness of GPT-UAI prior to the midterm through any channel (emails, peer diffusion, or the course platform). For RCT 2, D^2 equals one if student i reports exposure to the guidance messages (directly via emails, or through peer diffusion)

IV-LATE Effects. I estimate instrumental variables (IV-LATE) specifications using treatment assignment as an instrument for D^1 and D^2, respectively. The first- and second-stage equations are:

$$D_i^j = \pi_0^j + \pi_1^j R_i^j + \pi_2^j X_i + \lambda_s + \eta_i^j \tag{2}$$

$$Y_i^j = \delta_0^j + \delta_1^j \widehat{D}_i^j + \delta_2^j X_i + \lambda_s + u_i^j \tag{3}$$

where $\widehat{D}_i^j$ is the predicted value from the first stage for RCT j. δ_1^j identifies a local average treatment effect under the standard assumptions, which we discuss below. In RCT 1, it captures the causal effect of awareness of GPT-UAI among students whose awareness was increased by the encouragement emails. In RCT 2, it captures the causal effect of exposure to guidance among

students whose exposure was induced by the guidance emails.

Identification. Identification of the ITT effects follows from random assignment, as supported by balance across pre-treatment covariates. Identification of the IV-LATE effects relies on the standard assumptions (Imbens and Angrist, 1994; Vytlacil, 2002). Relevance is satisfied as treatment assignment significantly increases awareness in RCT 1 and guidance exposure in RCT 2, as shown in the first-stage results (Panel A of Table 2 and Table 3).

The exclusion restriction requires that treatment assignment affects outcomes only through the induced exposure. The intervention emails primarily provide information about the GPT-UAI or guidance on how to use it, and do not introduce new course content. While the guidance emails may directly influence study behavior, the IV estimates should therefore be interpreted as capturing the effect of exposure to this guidance package, rather than isolating a single behavioral channel.

Monotonicity is plausible given that the interventions consist of simple informational messages, making it unlikely that assignment reduces awareness or exposure.

The presence of spillovers implies that the standard no-interference assumption is not imposed in this setting. Both the reduced-form and IV estimates should therefore be interpreted in an environment with interference across students. The ITT estimates capture the effect of offering the intervention when information can diffuse through peer interactions, rather than a direct treatment effect under no interference. Similarly, the IV estimates identify local average effects of exposure induced by the intervention, but these effects are shaped by the equilibrium level of diffusion. Accordingly, the estimands should be interpreted as policy-relevant effects under realistic conditions of partial saturation, rather than as isolated direct effects.

Estimation Procedures. I include strata fixed effects and baseline covariates in all specifications. Given the randomized nature of our treatment assignments, this inclusion serves to improve precision but does not change the magnitude of our estimates. Standard errors are heteroskedasticity-robust and not clustered, as randomization occurs at the individual level.[10] To account for multiple hypothesis testing, I include adjusted p-values using the Romano and Wolf (2005) procedure within the family of academic performance outcomes.

[10] Clustering would be required under cluster-level randomization, for example if entire class sections were assigned to treatment or control (Abadie et al., 2023).

6 Results

6.1 RCT 1: Encouragement Effects

Adoption. We start documenting that the randomized encouragement increases reported awareness and usage of GPT-UAI, even in a context of substantial baseline exposure. Table 2 presents these results in Panel A, columns 1 to 3.

Assignment to receive the encouragement emails raises awareness by 21 percentage points (column 1) from 66% in the control group,[11] while usage goes up by 11 percentage points (column 2) over a 51% baseline. It also increases reported usage intensity by about 1.2 additional uses relative to a control mean of 1.8 (column 3), which is equivalent to a 68% increase.

Learning Outcomes. Despite these increases in awareness and usage, we find no corresponding effects on midterm performance, as shown in Panel B of Table 2.

The intention-to-treat (ITT) effects are positive, but small and not statistically distinguishable from zero at conventional levels across specifications in columns 1 to 3. The reduced-form effect on standardized grades is 0.05 standard deviations (SD), noisily measured. On the original 1–7 grading scale, the difference is 0.06, which is essentially a null effect; the corresponding means are 3.71 for the control group and 3.77 for the treatment group. The ITT effect on the raw score (0–100 scale) is 0.74 points favoring the treatment group. This effect is non-significant and very small compared to the control mean of 40.9 points.

Table 2 also reports the IV-LATE estimates instrumenting GPT awareness with randomization in Panel B (columns 4-6). Consistent with the ITT results, the estimates are noisy and statistically indistinguishable from zero for all three outcomes.[12]

Mechanisms: Adoption without Performance Effects. One potential mechanism behind these null effects is that students may not have perceived GPT-UAI as particularly useful, especially at a time when many were still experimenting with AI tools. Columns 4–6 in Table 2 provide

[11] This 66% can be decomposed into 52 percentage points (pp) who report discovering the GPT-UAI on the course platform, 8 pp through peers, 3 pp through other sources, and 3 pp through emails. The remaining 34% of the control group report not knowing about the GPT at midterm. Appendix Table E.1 reports the full distribution by treatment status.

[12] The results are qualitatively the same when instrumenting the variables of GPT-UAI usage (both at least once and number of times) with randomization, as we show in Table E.3.

evidence consistent with this interpretation. The average perceived usefulness in the control group is 3.7 on a 0–10 scale, and treatment effects on both this continuous measure and indicators of high usefulness are small and imprecise. The intervention increased awareness and usage but did not increase perceived value. Students may have viewed the GPT as one among several study inputs rather than as a particularly valuable resource.

In addition, the midterm contributes to the course grade but does not determine course completion or progression in the 3+2 program. Incentives to adjust study behavior may therefore be weaker at this stage of the semester relative to the final exam.[13]

Together, these results suggest that increasing adoption at the extensive margin, in a context of ongoing diffusion, is not sufficient to generate measurable learning gains on the midterm exam. This finding motivates the second intervention, which aims to influence usage quality rather than expand adoption.

[13] A complementary explanation is statistical and related to the power of the experiment. If true effects at the midterm are small, a larger sample could have detected them with greater precision.

Table 2: RCT 1 Estimation Results

Panel A: Adoption & Usage

	(1) GPT Awareness	(2) GPT Use	(3) Times Used	(4) Useful(0-10)	(5) Useful (7to10)	(6) Useful (=10)
RCT 1	0.213***	0.112*	1.214**	0.517	0.035	0.044
	(0.047)	(0.061)	(0.509)	(0.464)	(0.059)	(0.041)
Control Mean	0.658	0.505	1.772	3.668	0.332	0.099
Control SD	0.475	0.501	3.179	3.751	0.472	0.299
Observations	303	303	303	303	303	303

Panel B: Learning Outcomes - Midterm Exam

	Intention-to-Treat			IV-LATE		
	(1) Grade (SD)	(2) Raw Grade	(3) Score	(4) Grade (SD)	(5) Raw Grade	(6) Score
RCT 1	0.053	0.060	0.737			
	(0.116)	(0.132)	(2.099)			
	[0.647;0.674]	[0.647;0.674]	[0.726;0.732]			
GPT Awareness				0.250	0.283	3.462
				(0.537)	(0.607)	(9.669)
				[0.642;0.843]	[0.642;0.843]	[0.721;0.847]
Control Mean	-0.000	3.708	40.905	-0.000	3.708	40.905
Control SD	1.000	1.130	18.524	1.000	1.130	18.524
Observations	303	303	303	303	303	303

Notes: Table 2 reports estimates from equations 1, 2, and 3 for RCT 1. Panel A includes GPT Awareness (indicator for knowing about GPT-UAI), GPT Use (indicator for having used GPT-UAI at least once), Times Used (self-reported number of uses), Useful (perceived usefulness, 0–10 scale), Useful (7–10) (indicator for high usefulness), and Useful = 10 (indicator for highest usefulness). Panel B includes standardized grades (mean zero in the control group), raw grades (1–7 scale), and exam scores (0–100 scale). All specifications include stratum fixed effects and baseline covariates from Table 1. Robust standard errors are in parentheses. Brackets report the unadjusted p-values and Romano-Wolf adjusted p-values, respectively. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

6.2 RCT 2: Guidance Effects

Guidance Exposure and Usage. The randomized guidance intervention increases exposure to the guidance content and usage of GPT-UAI in a context of widespread generic AI adoption. Table 3 presents these results in Panel A, columns 1 to 4.

Assignment to receive the guidance emails raises reported exposure by 58 percentage points (column 1) relative to a control mean of 35 percent. This exposure in the control group arises from 17 percent report receiving the guidance from peers and 18 percent report direct receipt of the guidance emails. (see Appendix Table E.2).

The guidance intervention also increases usage of GPT-UAI, with effects that are similar in magnitude to those observed in RCT 1. Those assigned to treatment report an increase in the use of the GPT-UAI by 11 percentage points relative to a control mean of 49 percent (column 2). It also raises the number of reported uses by 1.2, from 2.8 to approximately 4.0 on average (column 3), equivalent to a 43% increase.

The assignment to treatment in RCT 2 increases usage of the GPT-UAI, but does not raise the overall usage of other generic AI tools (generic ChatGPT, Gemini, Claude, Copilot, etc.). Column 4 in Table 3 shows that , 95 percent of control group students report using at least one of these AI tools for studying, with no differences with the treatment group.[14] This result reflects a setting in which generic AI use is already nearly universal.

Learning Outcomes. Table 3 reports the intent-to-treat (ITT) effects on final exam outcomes in Panel B. Assignment to guidance increases final grades by 0.22 standard deviations, raw grades by 0.21 and raw scores by 3.5 points (columns 1–3). The probability of getting a passing grade in the exam increases by 12.8 percentage points relative to a control mean of 65 percent (column 4).

For the subsample of Business Administration students, for whom the exam is decomposed into open-response and multiple-choice components, we find that the effects are concentrated in the open-response section. Treatment assignment increases open-response scores by 4.9 points, while estimates for the multiple-choice section are small and statistically indistinguishable from zero (columns 5–6). This pattern is consistent with the design of the guidance intervention, which emphasized structured

[14] We also show that usage does not change significantly for each generic AI tool by treatment status (see Appendix Table E.4).

reasoning and step-by-step problem solving.

Panel C in Table 3 reports instrumental variables (IV-LATE) estimates, where reported exposure to guidance is instrumented with random assignment. The estimates show that exposure to guidance increases grades by 0.39 standard deviations, raw grades by 0.38, raw scores by 6.3 points, and increases the probability of passing by 22 percentage points. As in the ITT results, where measurable, the effects are driven by improvements in the open-response component (8.8 points), with no detectable effect on multiple-choice performance.

Overall, the evidence indicates that guidance on how to use the GPT-UAI improves performance on the high-stakes final exam, particularly in components that require structured problem solving.

Table 3: RCT 2 Estimation Results

Panel A: Usage & Guidance Exposure

	(1) Guidance Exp.	(2) GPT Use	(3) Times Used	(4) Any AI use
RCT 2	0.583***	0.109*	1.170**	-0.014
	(0.046)	(0.059)	(0.545)	(0.028)
Control Mean	0.354	0.487	2.861	0.949
Control SD	0.480	0.501	4.316	0.220
Observations	289	289	289	289

Panel B: Learning Outcomes - Final Exam (Intention-to-treat)

	(1) Grade (SD)	(2) Raw Grade	(3) Score	(4) Pass	(5) Open Response	(6) Multiple Choice
RCT 2	0.218**	0.210**	3.486**	0.128***	4.990***	0.351
	(0.100)	(0.096)	(1.600)	(0.049)	(1.645)	(0.837)
	[0.032;0.049]	[0.032;0.049]	[0.033;0.049]	[0.009;0.017]	[0.002;0.004]	[0.556;0.544]
Control Mean	0.000	4.256	54.244	0.646	34.699	15.722
Control SD	1.000	0.963	16.040	0.480	12.551	6.278
Observations	289	289	289	289	194	194

Panel C: Learning Outcomes - Final Exam (IV-LATE)

	(1) Grade (SD)	(2) Raw Grade	(3) Score	(4) Pass	(5) Open Response	(6) Multiple Choice
Guidance Exp.	0.392**	0.377**	6.257**	0.224***	8.836***	0.855
	(0.175)	(0.169)	(2.813)	(0.084)	(2.743)	(1.406)
	[0.0256;0.039]	[0.0256;0.039]	[0.0261;0.039]	[0.008;0.017]	[0.001;0.003]	[0.543;0.541]
Control Mean	0.000	4.256	54.244	0.646	34.699	15.722
Control SD	1.000	0.963	16.040	0.480	12.551	6.278
Observations	289	289	289	289	194	194

Notes: Table 3 reports the results from estimating equations 1, 2 and 3 for RCT 2. Panel A includes Guidance Exposure (indicator for receiving guidance directly or through peers), GPT Use (indicator for having used GPT-UAI at least once), Times Used (self-reported number of uses), and Any AI Use (indicator for using any AI tool). Panel B includes standardized grades (mean zero in the control group), raw grades (1–7 scale), an indicator for getting a passing grade (above 4 in raw scale), and exam scores (0–100 scale), as well as subscores for open-response and multiple-choice components where applicable. All specifications include stratum fixed effects and baseline covariates from Table 1. Robust standard errors are in parentheses. Brackets report the unadjusted p-values and Romano-Wolf adjusted p-values, respectively. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Mechanisms: Shifting Usage and Study Portfolio. We now use further survey measures exploring mechanisms behind the effects on the final exam performance. The evidence suggests that learning gains arise from improving the quality of interaction with the tool, through higher reported usefulness and tutor mode use. We also find that randomizing guidance content increased the perceived usefulness of GPT-UAI relative to the other (standard) study resources.

The higher perceived usefulness results are presented in columns 1-3 in Panel A in Table 4. Students randomized to receive the guidance messages report a 1.4 increase on the continuous 0-10 usefulness scale, which is equivalent to a 35% increase over the control mean of 3.95. Because the control-group standard deviation of the 0–10 usefulness scale is 3.73, the estimated effect in column (1) corresponds to a 0.38 standard deviation. Randomization to guidance also translates into large increases in the probability of reporting high usefulness (15 pp over 38% for reporting categories 7 to 10), and top usefulness (15 pp over 7% for reporting category 10 only).

The assignment to guidance also increases reported tutor-oriented use of the GPT-UAI. While we find no detectable effect when using the continuous 0–5 tutor scale (column 4 in Panel A of Table 4), there are relevant effects concentrated among those reporting high levels of tutor interaction. Treatment increases the probability of reporting high tutor-oriented use (categories 4 or 5) by 10 pp and the probability of reporting the highest category (5 only) by 12 pp (columns 5 and 6). This behavioral shift aligns with the pattern of performance gains, which are concentrated in open-response questions requiring structured reasoning rather than in multiple-choice components.

Finally, the randomizing guidance also increases the perceived usefulness of GPT-UAI relative to other study resources over the semester, measured on a 1 to 5 scale. Table 4, Panel B, shows that assignment to guidance increases the perceived usefulness of GPT-UAI by 0.62 points relative to a control mean of 2.57. This is equivalent to a 24% shift in that perception. In contrast, perceived usefulness of teaching assistant sessions, lectures, and problem set guides does not increase. The only additional statistically significant effects are smaller reductions in the perceived usefulness of lab sessions (-0.22) and course readings (-0.33), suggesting some substitution away from traditional materials.

The control means also provide a ranking of study inputs in the absence of guidance. Students rate problem set guides (4.37) and teaching assistant sessions (3.69) as the most useful resources, followed by lectures (3.48). GPT-UAI receives a lower average rating (2.57), similar to readings

(2.56) and above lab sessions (2.00). The guidance intervention improves the relative position of the AI assistant within this ranking, narrowing the gap with more traditional instructional supports.

Overall, the evidence suggests that guidance increased the perceived value of the GPT as part of students' study portfolio over the semester, rather than displacing core course components.

Table 4: RCT 2: Exploring Mechanisms

Panel A: Perceived Usefulness and Interaction

	(1) Useful(0-10)	(2) Useful (7to10)	(3) Useful (=10)	(4) Tutor (0-5)	(5) Tutor=4or5	(6) Tutor=5
RCT 2	1.413***	0.153***	0.152***	0.170	0.104*	0.117**
	(0.458)	(0.059)	(0.040)	(0.167)	(0.055)	(0.059)
Control Mean	3.946	0.376	0.067	3.718	0.624	0.342
Control SD	3.729	0.486	0.251	1.316	0.486	0.476
Observations	289	289	289	289	289	289

Panel B: During the semester, how helpful was each resource for your learning? (1-5 scale)

	(1) GPT	(2) TA Sessions	(3) Lab Sessions	(4) Lectures	(5) Readings	(6) PSets Guides
RCT 2	0.623***	-0.216	-0.224*	0.139	-0.330**	0.127
	(0.182)	(0.149)	(0.129)	(0.147)	(0.155)	(0.105)
Control Mean	2.570	3.692	1.986	3.477	2.563	4.362
Control SD	1.485	1.324	1.162	1.291	1.395	0.974
Observations	289	289	289	289	289	289

Notes: Table 4 reports estimates from equation 1 for RCT 2. Panel A includes perceived usefulness (0–10 scale), indicators for high usefulness (7–10) and highest usefulness (10), and measures of interaction with the GPT, including tutor-style use (0–5 scale), Tutor = 4 or 5, and Tutor = 5. Panel B reports perceived usefulness (1–5 scale) of different study inputs during the semester, including GPT-UAI, teaching assistant sessions, lab sessions, lectures, readings, and problem set guides. All specifications include stratum fixed effects and baseline covariates from Table 1. Robust standard errors are in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

7 Conclusions

This paper provides experimental evidence on the effects of a course-specific AI assistant in higher education, at a time when generative AI tools are rapidly diffusing across universities. I study two margins of AI use through two randomized interventions implemented in a real-world course setting. The design combines administrative and survey data to measure academic outcomes, usage, compliance, and spillovers, and evaluates behavior under real academic stakes.

Encouraging adoption increases awareness and usage of the GPT-UAI tool but does not improve academic performance. In contrast, providing guidance on how to use the tool leads to meaningful gains in final exam outcomes, with effects concentrated in tasks requiring structured reasoning. These results indicate that access alone is not sufficient, and that learning depends on how students use AI tools.

While specific to this setting, the findings inform a broader higher-education environment in which access to generative AI is widespread elsewhere. Therefore, relevant margin appears to be how the already in place AI is integrated into the learning process.

The interventions studied here are low-cost and operate within existing course infrastructure, making them informative for institutions considering how to respond to widespread AI use. The evidence suggests that policies aimed at guiding students toward more effective use of AI tools may be more consequential than policies focused solely on access or restrictions.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process During the preparation of this work the author(s) used ChatGPT and Claude in order to check the writing of the paper. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2023). When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, 138(1):1–35.

Agrawal, K., Athey, S., Kanodia, A., Nath, S., and Palikot, E. (2026). The economics of algorithmic personalization: Evidence from an educational technology platform. Working Paper 34950, National Bureau of Economic Research.

Amoah, A., Asiama, R. K., and Kwablah, E. (2025). Chatgpt early usage among students: A global evidence of determinants. *Development and Sustainability in Economics and Finance*, 7:100065.

Athey, S. and Imbens, G. W. (2017). The econometrics of randomized experiments. *Handbook of Economic Field Experiments*, 1:73–140.

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.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1(4):200–232.

Contractor, Z. and Reyes, G. (2025). Generative ai in higher education: Evidence from an elite college. IZA Discussion Paper 18055, Institute of Labor Economics (IZA), Bonn.

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., and Severín, E. (2017). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics*, 9(3):295–320.

Cueto, S., Beuermann, D. W., Cristia, J., Malamud, O., and Pardo, F. (2025). Laptops in the long run: Evidence from the one laptop per child program in rural peru. *Journal of Public Economics*, 252:105538.

de Barros, A. and Ganimian, A. J. (2024). Which students benefit from computer-based individualized instruction? experimental evidence from public schools in india. *Journal of Research on Educational Effectiveness*, 17(2):318–343.

De Simone, M., Tiberti, F., Rodriguez, M. B., Manolio, F., Mosuro, W., and Dikoru, E. J. (2025). From chalkboards to chatbots: Evaluating the impact of generative ai on learning outcomes in nigeria. Policy Research Working Paper 11125, World Bank, Washington, DC.

Duflo, E., Glennerster, R., and Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics*, 4:3895–3962.

Fischer, M., Rau, H. A., and Rilke, R. M. (2025). Ai tutoring enhances student learning without crowding out reading effort. Technical Report 18338, IZA Discussion Paper Series.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Jin, Y., Yan, L., Echeverria, V., Gašević, D., and Martinez-Maldonado, R. (2025). Generative ai in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, 8:100348.

Kestin, G., Miller, K., Klales, A., et al. (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.

Malamud, O. and Pop-Eleches, C. (2011). Home computer use and the development of human capital *. *The Quarterly Journal of Economics*, 126(2):987–1027.

Muralidharan, K., Singh, A., and Ganimian, A. J. (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review*, 109(4):1426–1460.

Nickow, A., Oreopoulos, P., and Quan, V. (2020). The impressive effects of tutoring on prek–12 learning: A systematic review and meta-analysis of the experimental evidence. *NBER Working Paper*, (27476).

Rodriguez-Segura, D. (2021). Edtech in developing countries: A review of the evidence. *The World Bank Research Observer*, 37(2):171–203.

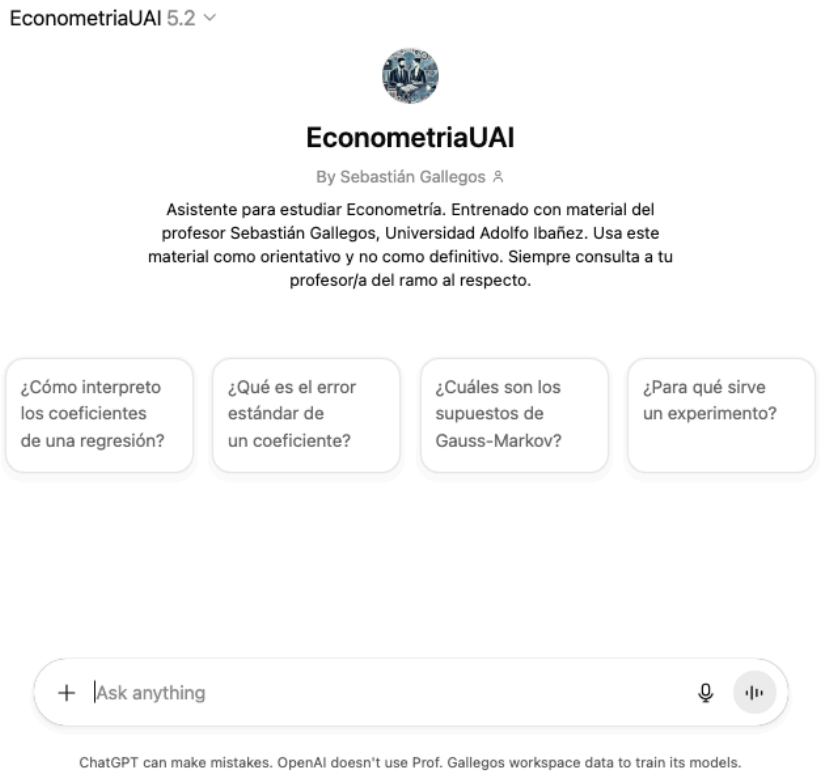
Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1):331–341.

Appendix

A AI Study Assistant (GPT-UAI) Interface

Figure A.1: Screenshot of the AI Study Assistant (GPT-UAI) Interface



Notes: Figure A.1 displays a screenshot of the AI study assistant (GPT-UAI) interface. The welcome message explains that EconometriaUAI is an assistant for studying Econometrics, trained with course materials at UAI. It states that responses should be considered indicative rather than definitive and that students should consult the course instructor when using the tool.

B Course Syllabus and Content

Original Version (Spanish)

Unidad 1: Introducción y repaso estadístico-matemático

1. Revisión del programa, marco general y presentación del curso
2. ¿Qué es la Econometría? Conceptos básicos. Predicción y causalidad
3. Revisión de probabilidad y estadística. Muestreo simple. Parámetros muestrales
4. Repaso de álgebra matricial

Lecturas sugeridas:

- Stock y Watson (2009), Caps. 1, 2 y 3
- Wooldridge (2010), Cap. 1; Apéndices A, B y C

Unidad 2: Modelo de regresión lineal

1. Derivación, intuición, estimación e interpretación del método de Mínimos Cuadrados Ordinarios (MCO)
2. Propiedades mecánicas de MCO
3. Propiedades estadísticas de los estimadores MCO bajo supuestos clásicos
4. Inferencia: estimación por intervalo y test de hipótesis
5. Medidas de bondad de ajuste
6. Predicción y análisis de residuos

Lecturas sugeridas:

- Stock y Watson (2009), Caps. 4–9
- Wooldridge (2010), Apéndice D; Caps. 2, 3, 4, 6, 7 (sec. 7.1–7.4), 8 y 9

Unidad 3: Causalidad y diseño en análisis empírico

1. Causalidad versus correlación, contrafactuales
2. Experimentos aleatorios controlados (RCTs) y aplicaciones
3. Validez externa vs validez interna
4. Omisión de variable relevante e inclusión de variable intrusa
5. Formas funcionales, elección de regresores e interpretación de coeficientes
6. Multicolinealidad
7. Heterocedasticidad

Lecturas sugeridas:

- Stock y Watson (2009), Caps. 6–9
- Wooldridge (2010), Apéndice D; Caps. 3, 4, 6, 7 (sec. 7.1–7.4), 8 y 9

Unidad 4: Modelos de elección discreta

1. Modelo de probabilidad lineal
2. Modelos Logit y Probit
3. Interpretación
4. Predicción y aplicaciones prácticas

Lecturas sugeridas:

- Wooldridge (2010), Cap. 7 (sec. 7.5–7.6) y 17

Unidad 5: Introducción a machine learning

1. Big Data y análisis económico
2. Machine learning: conceptos y aplicaciones
3. Machine learning versus econometría
4. Algoritmos más habituales en machine learning

Lecturas sugeridas:

- Mullainathan and Spiess (2017), *Journal of Economic Perspectives*

- Athey and Imbens (2019), *Annual Review of Economics*

Translated Version (English)

Unit 1: Introduction and statistical-mathematical review

1. Review of the syllabus, general framework, and course presentation
2. What is Econometrics? Basic concepts: prediction and causality
3. Review of probability and statistics. Simple sampling. Sample parameters
4. Review of matrix algebra

Suggested readings:

- Stock and Watson (2009), Chapters 1–3
- Wooldridge (2010), Chapter 1; Appendices A–C

Unit 2: The linear regression model

1. Derivation, intuition, estimation, and interpretation of Ordinary Least Squares (OLS)
2. Mechanical properties of OLS
3. Statistical properties of OLS estimators under classical assumptions
4. Inference: confidence intervals and hypothesis testing
5. Measures of goodness of fit
6. Prediction and residual analysis

Suggested readings:

- Stock and Watson (2009), Chapters 4–9
- Wooldridge (2010), Appendix D; Chapters 2–4, 6–9

Unit 3: Causality and empirical research design

1. Causality versus correlation, counterfactuals
2. Randomized controlled trials (RCTs) and applications
3. External validity versus internal validity
4. Omitted variable bias and irrelevant regressors
5. Functional form, choice of regressors, and interpretation of coefficients
6. Multicollinearity
7. Heteroskedasticity

Suggested readings:

- Stock and Watson (2009), Chapters 6–9
- Wooldridge (2010), Appendix D; Chapters 3–4, 6–9

Unit 4: Discrete choice models

1. Linear probability model
2. Logit and Probit models
3. Interpretation
4. Prediction and practical applications

Suggested readings:

- Wooldridge (2010), Chapter 7 (Sections 7.5–7.6) and Chapter 17

Unit 5: Introduction to machine learning

1. Big data and economic analysis
2. Machine learning: concepts and applications
3. Machine learning versus econometrics
4. Common machine learning algorithms

Suggested readings:

- Mullainathan and Spiess (2017), *Journal of Economic Perspectives*
- Athey and Imbens (2019), *Annual Review of Economics*

C Email Messages – RCT 1 (Encouragement Intervention)

Email 1 — Monday at 12:00 PM (Two days before the midterm)

Subject: Have you tried the EconometríaUAI GPT?

Dear [NAME],
Remember that you have access to **GPT EconometríaUAI**, a tool specially trained with course materials to help you study.
⇒ **Try it here:**
https://chatgpt.com/g/g-ftvZAGEWO-econometriauai

Explore it today and get better prepared for Wednesday's exam!

Sincerely,
Econometrics Coordination

Email 2 — Tuesday at 11:00 AM (One day before the midterm)

Subject: Questions about Midterm 1? GPT can help you

Dear [NAME],
It is normal to have questions the day before an exam. GPT EconometríaUAI can support you. Take advantage of this tool to review your knowledge.

For example, ask:

- How do I interpret the constant in a simple regression?

GPT response: The constant (intercept) in a simple linear regression represents the predicted value of the dependent variable when the explanatory variable equals zero.

⇒ **Try it now with your own questions:**
https://chatgpt.com/g/g-ftvZAGEWO-econometriauai

Sincerely,
Econometrics Coordination

Email 3 — Wednesday at 10:00 AM (Morning of the midterm)

Subject: Last questions before today's exam?

Dear [NAME],
Today at 6:00 PM is Midterm 1.
Do you have last-minute questions? Watch this short video (less than one minute) and learn how GPT can help you review:

⇒ **Watch video:**
https://www.youtube.com/watch?v=UNHFs4TKi5M

⇒ **Access GPT here:**
https://chatgpt.com/g/g-ftvZAGEWO-econometriauai

Best of luck,
Econometrics Coordination

D Email Messages – RCT 2 (Guidance Intervention)

Email 1 — Two days before the final exam (Morning)

Subject: GPT EconometríaUAI: Use it as a tutor

Dear [NAME],
Use **GPT EconometríaUAI** (link: https://chatgpt.com/g/g-ftvZAGEWO-econometriauai) in tutor mode and improve your learning.
Copy and paste the prompt below when studying for your exam:

> "Act as an econometrics tutor for my exam. Do not give me the final answer. Ask me three diagnostic questions (one on binary dependent variables, one on the Gauss–Markov assumptions, and one of your choice), then guide me step by step, asking me to complete the next step myself. At the end, give me a similar exercise to solve on my own and correct it."

Sincerely,
Econometrics Coordination

Email 2 — Two days before the final exam (Afternoon)

Subject: Practice exam questions + correction with GPT EconometríaUAI

Hello,
Here is a new suggestion to improve your exam preparation. Use practice and correction in **GPT EconometríaUAI** (link: https://chatgpt.com/g/g-ftvZAGEWO-econometriauai). Copy and paste the following prompt:

> "Generate five exam-style exercise questions (assumptions, interpretation, Probit, Logit, and linear probability model). Do not provide solutions. I will answer them one by one: correct me and explain any mistakes in two lines. Then give me a similar variation so I can practice again."

Good luck with your studying,
Econometrics Coordination

Email 3 — One day before the final exam (Afternoon)

Subject: GPT EconometríaUAI – Verify responses when studying

Dear [NAME],
It's important to avoid mistakes or overconfidence. Verify **GPT EconometríaUAI's** (link: https://chatgpt.com/g/g-ftvZAGEWO-econometriauai) answers by using the prompt below:

> "Present and solve an exam-style exercise. Then verify the answer you just gave me: (1) state the necessary assumptions, (2) list two common mistakes a student might make, (3) provide a quick sense check (sign and magnitude), and (4) indicate which part could be incorrect or ambiguous."

Good luck on tomorrow's exam,
Econometrics Coordination

E Additional Tables

Table E.1: How Students Learned About GPT-UAI (RCT 1)

Aware of the GPT-UAI?	Randomized in RCT 1 to		
	Control	Treatment	Total
I did not know it existed	34.16	11.88	26.73
Yes, I know the GPT	65.84	88.12	73.27
Total	100.00	100.00	100.00
Number of Observations	202	101	303

Source of Information	Control	Treatment	Total
(a) Encouragement email	2.97	68.32	24.92
(b) Classmate or peer	7.92	1.98	5.90
(c) Found it on the course platform	51.98	15.84	40.00
(d) Other source	2.97	1.98	2.62
(e) I did not know it existed	34.16	11.88	26.56
Total	100.00	100.00	100.00
Number of Observations	202	101	303

Notes: Table E.1 reports the percentage distribution of students by how they report they first learned about GPT-UAI in RCT 1, separately for the control and treatment groups in the midterm encouragement experiment.

Table E.2: Receipt and Sharing of Guidance Emails (RCT 2)

	Randomized in RCT 2 to		
	Control	Treatment	Total
No, but someone shared the advice with me	16.78	5.71	11.42
No, and no one shared the advice with me	65.77	7.86	37.72
Yes, and I shared the advice with others	6.04	19.29	12.46
Yes, but I did not share the advice	11.41	67.14	38.41
Total	100.00	100.00	100.00
Number of Observations	149	140	289

Notes: Table E.2 reports the percentage distribution of students by whether they received the guidance emails in RCT 2 and whether they shared the advice with others. Percentages are shown separately for control and treatment groups.

Table E.3: IVs (RCT 1)

	(1) Grade (SD)	(2) Raw Grade	(3) Score	(4) Grade (SD)	(5) Raw Grade	(6) Score
GPT Use	0.477	0.538	6.591			
	(1.044)	(1.179)	(18.562)			
Times Used				0.044	0.050	0.606
				(0.094)	(0.106)	(1.690)
ControlMean	-0.000	3.708	40.905	-0.000	3.708	40.905
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Observations	303	303	303	303	303	303

Notes: This table reports IV-LATE estimates instrumenting the variables of GPT usage (both at least once and number of times) with randomization. Robust (Huber-White) standard errors in parentheses.

Table E.4: Usage of other AI tools (RCT 2)

	(1) ChatGPT Paid	(2) ChatGPT Free	(3) Gemini Paid	(4) Gemini Free	(5) Copilot	(6) Claude Free	(7) Claude Paid	(8) Other IA
RCT 2	0.071	-0.045	0.013	-0.008	-0.026	-0.013	-0.019*	-0.028
	(0.058)	(0.058)	(0.056)	(0.045)	(0.028)	(0.025)	(0.011)	(0.024)
ControlMean	0.430	0.403	0.329	0.174	0.067	0.054	0.020	0.067
ControlSD	0.497	0.492	0.471	0.381	0.251	0.226	0.141	0.251
Observations	289	289	289	289	289	289	289	289

Notes: Table E.4 reports the results from estimating equation 1 for RCT 2 on the respective dependent variables. These correspond to indicator variables for whether students report using each generic AI tool. Robust (Huber-White) standard errors in parentheses.

F No Differential Participation by Treatment Status

Table F.5: Participation in Analytical Samples by Treatment Status

	(1) Sample RCT 1	(2) Sample RCT 1	(3) Sample RCT 1	(4) Sample RCT 2	(5) Sample RCT 2	(6) Sample RCT 2
RCT 1	0.009	0.007	0.020			
	(0.044)	(0.043)	(0.043)			
RCT 2				-0.008	-0.012	-0.010
				(0.044)	(0.043)	(0.043)
ControlMean	0.789	0.789	0.789	0.753	0.753	0.753
Strata	No	Yes	Yes	No	Yes	Yes
Covariates	No	No	Yes	No	No	Yes
Observations	384	384	384	384	384	384

Notes: Table F.5 reports a host of estimations showing that participation in the RCTs is balanced by treatment status. Columns 1 to 3 report regressions on an indicator for participation in the analytical sample of RCT 1 using the all 384 students enrolled at the beginning of the semester. The first regression does not include strata or covariates, the second adds strata, and the third adds covariates. Columns 4 to 6 report analogous regressions for RCT 2. Robust (Huber-White) standard errors in parentheses.

Table F.6: Mean Characteristics for the Randomization Sample and Analytical RCTs Samples

	(1) Randomization Sample	(2) Analytical Sample RCT 1	(3) Sample RCT 2
Randomized to Treatment in RCT 1	0.331	0.333	0.332
Randomized to Treatment in RCT 2	0.487	0.492	0.484
Female	0.30	0.31	0.31
Private High School	0.75	0.76	0.74
Degree Progression	67.70	66.97	67.32
Major Business	0.72	0.70	0.67
Entrance Cohort	2022.9	2022.9	2022.9
Exact Age	21.90	21.80	21.82
Birth Year	2003.29	2003.39	2003.38
Birth Month	6.74	6.74	6.65
Birth Day	15.37	15.58	15.59
Section 1	0.141	0.152	0.152
Section 2	0.164	0.129	0.173
Section 3	0.154	0.165	0.138
Section 4	0.104	0.119	0.104
Section 5	0.156	0.135	0.104
Section 6	0.161	0.168	0.194
Section 7	0.120	0.132	0.135
Observations	384	303	289
p-value of joint test	-	0.658	0.599

Notes: Table F.6 reports mean characteristics for the randomization sample (column 1) and the analytical RCTs samples (in columns 2 and 3). The bottom panel reports the number of observations and the p-value from a joint test of the baseline covariates for each analytic sample.