Tackling Biased Beliefs to Foster Women's Entry into Economics: Experimental Evidence from Mexico*

Stanislao Maldonado ®
Tecnológico de Monterrey

Edgar Castro Méndez ®
Tecnológico de Monterrey

Elvira Naranjo ®
Tecnológico de Monterrey

Joana Chapa ®
Universidad Autónoma de Nuevo León

This version: October 31, 2026

Abstract

We study whether a brief exposure to female role models can change biased beliefs about economics and close the gender gap in choosing economics as a university major. We conducted a field experiment in Nuevo León, Mexico, in which female economists delivered a 35–45 minute information-and-narrative session. The intervention broadened students' perceptions of economics beyond money and finance. An index capturing reported understanding of economists' work, the field's perceived social value, and rejection of this stereotype increased by 0.30 standard deviations in the short term and by 0.17 at follow-up, relative to the control group. However, the estimated effect on women's applications to economics was statistically indistinguishable from zero. We also find a statistically significant increase of 0.66 percentage points in applications to UANL Economics in the pooled sample of final-year students from both school networks. Overall, the intervention broadened students' perceptions of economics, with effects persisting at follow-up, but we find no statistically significant effect on women's applications to economics programs.

Keywords: Gender gaps; role models; motivated beliefs; economics education; randomized controlled trial; Mexico.

JEL Classification: C93, I23, J16, D90.

*Emails: stanislao.maldonado@tec.mx (corresponding author); ecastrom@tec.mx; enaranjo@tec.mx; joana.chapacn@uanl.edu.mx. We owe particular thanks to Héctor García and Lianet Farfán Pérez, who coordinated field implementation. We are also grateful to the economists who volunteered as role models and to the school principals and teachers in the UANL and Prepa Tec networks for facilitating the study. We thank the research assistants who supported fieldwork and data collection. The International Economic Association provided financial support for the fieldwork. The study was approved by the Research Ethics Committee under protocol P-ECSG-202511-01. This randomized controlled trial is registered in the AEA Social Science Registry under ID AEARCTR-00017053; the pre-analysis plan is available at https://www.socialscienceregistry.org/trials/17053. The order of authors was randomized using the Author Randomization Tool of the American Economic Association.

1 Introduction

Women remain substantially underrepresented in economics despite broad gains in women's access to higher education and to quantitatively oriented fields in recent years (Kahn, 1995; Dynan and Rouse, 1997; Ginther and Kahn, 2004). The gap is visible when students choose a major and becomes more pronounced along the academic pipeline, where women account for progressively smaller shares of senior faculty, research leaders, and other positions that shape the questions the discipline asks (Kahn, 1993, 1995; Ginther and Kahn, 2004). This imbalance matters beyond the composition of economics departments. Economists influence the design of public policy, the allocation of public resources, and the evidence used to evaluate social programs (Christensen, 2018; Christensen and Mandelkern, 2022). A profession that draws from a narrower pool of experiences may therefore also narrow the set of questions, mechanisms, and policy tradeoffs that receive sustained attention (Bayer and Rouse, 2016; Ginther, Kahn and McCloskey, 2016).[1]

Part of this gap may originate well before students enter a university classroom (Bertocchi, Bonacini and Murat, 2023). Economics is not usually taught as a distinct field in upper-secondary school, and students often encounter it through a narrow set of visible topics—inflation, finance, business, or exchange rates (Dynan and Rouse, 1997; Bansak and Starr, 2010; Owen and Hagstrom, 2021). This exposure can obscure the discipline's diversity in research topics, including issues such as poverty, inequality, health, education, and other problems that many students, particularly female students, regard as socially consequential (Buzard et al., 2025; Bansak and Starr, 2010; Owen and Hagstrom, 2021). It can also make economics appear more mathematically forbidding and less compatible with an identity centered on helping others (Bansak and Starr, 2010; Steele, 1997; Barrera et al., 2024).

These perceptions need not be accurate to affect choices. When students have little direct information about a field, their application decisions may respond to subjective beliefs about what economists do, who belongs in the profession, and whether their own skills and preferences fit that image (Dynan and Rouse, 1997; Bansak and Starr, 2010; Ersoy and Speer, 2025). Gender differences in quantitative self-assessment and perceived belonging may amplify this process even

[1] A large literature documents gender disparities in entry into economics and progression within the profession. Bayer and Rouse (2016), Lundberg and Stearns (2019), and Buckles (2019) provide broad accounts of the pipeline in the United States, while Avilova and Goldin (2024) discuss barriers that operate before and during the choice of economics. WELAC (2023) and Arceo-Gómez (2025) document related patterns in Latin America and Mexico.

when academic preparation is similar (Buzard et al., 2025; Steele, 1997; Owen, 2010).[2]

These barriers have motivated efforts to increase women's participation in economics through information provision, encouragement, mentoring, curricular changes, and exposure to female role models (Avilova and Goldin, 2024; Barrera et al., 2024). Role-model interventions are particularly relevant because they can address several of these barriers within a single encounter. A female economist can provide information about the discipline, illustrate a possible career path, and challenge perceptions of who belongs in the profession. Exposure to female role models has increased women's participation in economics and selected STEM programs (Porter and Serra, 2020; Breda et al., 2023; Agurto et al., forthcoming). Yet students may revise their understanding of a field, their expectations of success, or their perceptions of its compatibility with their interests, and these responses need not occur together. Understanding which dimensions change, whether those changes persist, and whether they translate into educational choices is therefore central to evaluating these interventions.

This paper asks whether a brief talk by a female economist can change the beliefs and choices underlying this gap. We examine whether the intervention changes students' understanding of economics and perceptions of fit, whether women respond differently, whether any effects persist, and whether they extend to applying to study economics.[3]

Mexico provides a particularly useful setting in which to study these questions. Women's university enrollment has expanded, but their representation in economics has remained comparatively stable. Women accounted for about 41 percent of Mexican undergraduate economics graduates in the 2019 WELAC survey and for roughly 32 percent of economics majors in our setting. Their representation declines further at later stages of the profession (WELAC, 2023; Arceo-Gómez, 2025). These patterns arise in a higher-education system that contains large public universities as well as selective private institutions and in which many students choose a field near the end of upper-secondary school. Nuevo León allows us to observe this decision in two contrasting university

[2]The relevant distinction is between ability and beliefs about ability or fit. The intervention cannot alter students' accumulated mathematics preparation in a single classroom period, but it may alter how they interpret that preparation and how they map their interests into a field of study. This distinction motivates our separate measures of beliefs about economics, identity and self-concept, intentions, and gender stereotypes; see also Barrera et al. (2024) and Avilova and Goldin (2024).

[3]An intervention may correct a misperception without changing a student's preferred major, and an immediate response may fade before an application is submitted. We therefore treat belief updating, persistence, and applications as distinct empirical margins (Pugatch and Schroeder, 2021; Ersoy and Speer, 2025; Patnaik et al., 2024; Agurto et al., forthcoming).

networks: the public Universidad Autónoma de Nuevo León (UANL) and the private Tecnológico de Monterrey (Tec).[4]

We partnered with upper-secondary schools affiliated with UANL and Tec and implemented an in-person role-model intervention during the 2025/26 academic year. Trained female professionals delivered a 35–45-minute session during a compulsory career-planning course. The session combined information about the breadth and social relevance of economics and labor market opportunities with exercises challenging common stereotypes and the economists' personal accounts of entering a historically male-dominated field. Its length and placement made it compatible with the career guidance that schools already provide. To estimate the impact of role models, we compared this intervention with an active placebo of the same general structure and duration: a university student discussed career choice, presented a personal narrative and videos, and invited questions, without promoting economics or discussing gender. Thus, the comparison holds constant the arrival of an outside presenter, the salience of an impending career decision, and much of the format of the classroom interaction.[5]

The evaluation uses a two-stage cluster-randomized design. The first-stage frame contained 29 school randomization units spanning 30 campuses. We assigned these units to treatment or control, stratifying by university network; 28 units spanning 29 campuses contribute student data to the analysis. We randomly selected six classrooms in each school. Within each treated unit, we then randomly selected half of the participating classrooms to receive the role-model session. The remaining classrooms received the active placebo within the same schools. This structure produces three groups: classrooms assigned to the role-model session in treated schools, classrooms assigned to the placebo in treated schools (the "spillover" group), and classrooms assigned to the placebo in control schools. The first-stage assignment identifies the main school-level intention-to-treat effect; the second assignment reveals whether responses are concentrated among directly exposed students or diffuse to classmates elsewhere in the same school.

[4]The 41 percent figure refers to graduates of undergraduate programs in the Mexican institutions covered by the 2019 WELAC survey, whereas the 32 percent figure refers to enrolled economics majors at UANL and Tec, according to administrative records.

[5]The treatment is a bundle. Relative to the active placebo, it varies the content of the session, the presenter's field, and the presenter's career stage, while making a successful woman in economics salient. Ideally, one might cross-randomize each component to distinguish information, role-model identity, seniority, and narrative effects (Serra, 2025; Agurto et al., forthcoming; Di, 2025). The available design instead identifies their combined reduced-form effect. We therefore use "role-model intervention" as a description of the bundle and do not attribute the estimates to a single psychological channel.

Across the 28 contributing school units, 3,763 students completed the baseline in 159 classrooms. We administered baseline and immediate post-session surveys, followed by a follow-up survey several weeks later. Through these surveys, we collected information on students' beliefs about economics, identity and self-concept, intentions and aspirations, and gender stereotypes. These four outcome families were pre-specified in the pre-analysis plan (PAP). After the admissions cycle, we fielded an online endline and linked the sample to the applicant registries of UANL and Tec. The pre-registered primary outcome is an application to economics among female final-semester students.

The main result is that the intervention broadens students' perceptions of economics beyond money and finance, both immediately after the session and at follow-up. Assignment to a treatment school increases the immediate beliefs-about-economics index by 0.30 standard deviations with respect to the control group. The standard error (S.E.) is 0.06. The immediate identity-fit measure also increases, by 0.10 standard deviations (S.E. = 0.03). Both effects are statistically significant. The beliefs result is robust to alternative inference procedures: the wild cluster bootstrap and randomization inference yield p-value < 0.001, as does conventional school-clustered inference. Both effects also remain significant after the four-domain Benjamini–Hochberg adjustment (Benjamini and Hochberg, 1995). On the other hand, intentions and gender stereotypes do not change detectably.

At follow-up, the school-level effect on beliefs about economics remains positive and statistically significant at 0.17 standard deviations (S.E. = 0.05). Alternative inference procedures preserve this conclusion (wild cluster bootstrap p-value = 0.009; randomization inference p-value = 0.008), as does adjustment for multiple testing. We find no evidence of effects on identity and self-concept, intentions and aspirations, or gender stereotypes.[6]

Although gender is central to the intervention, the belief response is not confined to women. The immediate estimate is 0.27 standard deviations for men and 0.34 for women. The women-minus-men difference is 0.07 standard deviations and is statistically indistinguishable from zero (S.E. = 0.07). The positive effects for both groups are consistent with information about economics being relevant to students beyond the population whose representation motivated the intervention.

[6]The immediate survey was administered minutes after the presentation, leaving experimenter demand as a possible contributor to the initial response (de Quidt, Haushofer and Roth, 2018). The later belief effect provides evidence beyond that setting, although incomplete follow-up and differences in measurement timing across arms limit conclusions about the duration of the response.

The two-stage assignment clarifies where the change occurs. Relative to control classrooms, immediate beliefs increase by 0.58 standard deviations in classrooms receiving the role-model session, while the estimated spillover effect in treated schools is close to zero, at 0.04 standard deviations. The incremental effect of the role-model session within a treated school is 0.54 standard deviations. At follow-up, the direct-treatment-versus-control effect is 0.31 standard deviations, while the estimated spillover effect is close to zero, at 0.05 standard deviations. We find no evidence of spillover effects on beliefs about economics, either immediately after the session or at follow-up. Thus, the evidence indicates belief updating beyond the immediate session, concentrated among students in direct-treatment classrooms.

For women's applications to economics, we find no evidence of an effect in our setting. The school-level estimate is 0.40 percentage points (p.p.) (S.E. = 0.27), relative to a control mean of 0.20 percent. It remains statistically indistinguishable from zero under conventional inference. This conclusion is unchanged under the wild cluster bootstrap (p-value = 0.182) and randomization inference (p-value = 0.211). We likewise find no statistically significant effect on the application outcome in the full sample. However, we do find a statistically significant increase of 0.66 p.p. in applications to UANL Economics in the pooled sample of final-year students from both school networks.

The paper makes two contributions. First, it adds to the experimental literature on role models and field-of-study choices. We examine belief updating and educational choices before university entry, complementing studies of role models among university students and recent interventions aimed at expanding interest in economics during high school (Porter and Serra, 2020; Patnaik et al., 2024; Gentry, Meer and Serra, 2023; Smith, Advani and Harvey, 2025). We compare the intervention with an active career-guidance session and follow students from immediate changes in beliefs and self-concept to follow-up responses and subsequent application decisions. These measurement horizons allow us to examine whether changes in how students view economics persist and translate into choices. Our findings show that a brief encounter with female economists can change perceptions of the discipline both immediately and at follow-up without a corresponding detectable increase in women's applications.[7]

[7]Information and encouragement have sometimes increased participation without attracting more women (Pugatch and Schroeder, 2021; Halim, Powers and Thornton, 2022). Related field experiments examine role models and social identity in technology and scalable role-model exposure in STEM and entrepreneurship (Del Carpio and

Second, the paper connects role-model interventions to the literature on motivated beliefs (Bénabou and Tirole, 2002, 2016). Recent studies have examined how the messages conveyed and the characteristics of the role models shape students' responses (Breda et al., 2023; Di, 2025; Serra, 2025). Our conceptual framework considers both learning about economics and the possibility that students' interpretation of information depends on its compatibility with their self-image. It distinguishes beliefs about the field, expectations of success, and the perceived value and costs of pursuing economics. This distinction is useful because students may acquire a broader understanding of the discipline without changing their preferred career. Our empirical analysis follows these different dimensions through measures of beliefs about economics, identity and self-concept, intentions and aspirations, and gender stereotypes, collected immediately after the intervention and at follow-up.[8]

More broadly, the results speak to the global effort to make economics a more diverse profession. The Undergraduate Women in Economics Challenge led by Claudia Goldin has evaluated a portfolio of light-touch strategies across U.S. colleges (Avilova and Goldin, 2018, 2024). Experiments generated by or closely related to this agenda report a mixed record: some find gains in course-taking or major choice from mentoring, targeted information, performance feedback, or female role models (Li, 2018; Bayer, Bhanot and Lozano, 2019; Porter and Serra, 2020; Bedard, Dodd and Lundberg, 2021; Chambers et al., 2021; Owen and Hagstrom, 2021; Avery et al., 2024; Patnaik et al., 2024), whereas others find no effects for women or only imprecise changes (Pugatch and Schroeder, 2021; Halim, Powers and Thornton, 2022; Antman, Skoy and Flores, 2026).

The remainder of the paper is organized as follows. Section 2 presents the conceptual framework. Sections 3 and 4 describe the institutional setting, intervention, and experimental design. Sections 5 and 6 present the data and empirical strategy. Section 7 reports the main results, and Sections 8 and 9 examine mechanisms and heterogeneity. Section 10 presents robustness exercises, Section 11 discusses cost-effectiveness, and Section 12 concludes.

Guadalupe, 2022; Asanov et al., 2026). These comparisons are informative about the potential reach of role models, but remain suggestive because the students, treatments, and decision margins differ.

[8]Related studies examine different dimensions of this decision. Agurto et al. (2025) study signals of mathematical aptitude, while Ersoy and Speer (2025) examine information about the financial and nonfinancial attributes of majors. Borges and Estevan (2024) study sustained exposure to female professors and peers during university. These contributions motivate distinguishing students' assessments of their own prospects from their understanding of a field and the environment in which they expect to study or work.

2 Conceptual Framework

We propose a simple framework to study students' field choice with incomplete information about both the field and their own prospects within it. In this framework, students must assess what economists do, whether they would succeed, whether the field fits their interests, and whether its expected benefits justify the academic, financial, and social costs of applying. A student can therefore be well prepared for economics and still decide against it because she holds an inaccurate belief about the discipline or about her place within it.

We use motivated beliefs to organize how the role model intervention may affect this decision (Bénabou and Tirole, 2002, 2016; Golman, Hagmann and Loewenstein, 2017). New information need not be processed mechanically: its influence may depend on whether the messenger is credible and compatible with the student's self-image and social environment. A female economist can provide first-hand information about the field while demonstrating that women enter and succeed in it.[9]

Let b_i denote student i's belief about her analytical ability, $p(b_i)$ her perceived probability of succeeding in economics, V_i the perceived value of the field, and C_i the perceived cost of pursuing it. Her perceived net return and application decision are

$$M_i = p(b_i)V_i - C_i, \qquad d_i = \mathbf{1}\{M_i \geq 0\}. \tag{1}$$

The treatment can operate through three channels. The role model may change beliefs about expected success, increase the perceived value of economics by emphasizing its career opportunities and social relevance, or reduce perceived barriers and lack of belonging. If the post-treatment values are (b_i', V_i', C_i'), the change in perceived net returns is

$$\Delta M_i = \left[p(b_i') - p(b_i)\right] V_i + p(b_i')\left(V_i' - V_i\right) - \left(C_i' - C_i\right). \tag{2}$$

The intervention is designed to increase the first two components and reduce the third. However, these changes are not mechanical. Students may begin with favorable priors, may not view the

[9] Throughout the paper, *identity alignment* means the perceived compatibility between economics and a student's interests, values, anticipated belonging, and self-concept.

presenter as a relevant comparison, or may update in an unintended direction. Moreover, a positive ΔM_i changes an application only for students sufficiently close to the threshold in Equation (1).

The framework yields three empirical predictions. First, students in treated schools—and, absent complete diffusion, especially those directly exposed—should report broader and more favorable beliefs about economics. Confidence, identity alignment, intentions, and stereotypes should move only if the session changes their corresponding components of b_i, V_i, or C_i. Second, these changes should remain at the follow-up only if the signal produces durable updating. Third, female final-semester students assigned to treated schools should be more likely to apply to economics only if the change in perceived net returns is sufficiently large to move them across the application threshold.

Finally, it is important to notice that the model does not unconditionally predict larger effects for women or for students with less favorable baseline beliefs. These groups may have more scope to update, but they may also lie farther from the application threshold or place less weight on the signal. We therefore evaluate heterogeneity with direct interaction tests. The second randomization stage provides a separate test of diffusion: spillover classrooms in treated schools respond only if students or teachers transmit some of the intervention's content.

These predictions organize the empirical analysis but do not distinguish motivated processing from standard information updating. A full version of this framework is available in the Online Appendix.

3 Institutional Background and Setting

3.1 The Mexican education system

Mexico's education system is organized into basic, upper-secondary, and higher education. The study takes place in upper-secondary education (*educación media superior*), which students normally enter after completing lower-secondary school. Programs at this level are commonly called *bachillerato* or *preparatoria*; we use "upper-secondary school" and "high school" interchangeably.[10]

[10]The Secretaría de Educación Pública (SEP) classifies preschool, primary, and lower-secondary schooling as basic education and places the typical upper-secondary age range at 15–17. Upper-secondary education includes the *bachillerato*, the professional-technical baccalaureate, and equivalent programs (Secretaría de Educación Pública, 2024). Because English-language education systems use "secondary school" differently, retaining the official Mexican term avoids conflating lower-secondary school with the population studied here.

Upper-secondary education became constitutionally compulsory in 2012. It includes academically oriented general baccalaureates and bivalent technological or professional-technical programs that combine preparation for higher education with occupational training. The distinction matters because curricula and contact with quantitative or social-science subjects differ across institutions, even though these routes can lead to university study.[11]

The national system does not impose a single upper-secondary sequence on the institutions in our study. General UANL *preparatorias* commonly operate four-semester programs, whereas PrepaTec and UANL's technical and bilingual schools operate six-semester programs. We therefore define final-semester students according to the last semester of their program rather than a common grade label.

Students near the end of either sequence must choose both a university and a field of study. Their exposure to economics before that choice depends on the curriculum and career guidance offered by their school. Panel A of Figure 1 locates the study population in the Mexican education system and maps the application, admission, and enrollment pathways observed in the UANL and Tec records. Panel B summarizes the program coverage and data sources used to measure these educational choices.

3.2 Female under-representation in economics in Mexico

Women remain underrepresented in economics in Mexico, although the available evidence does not describe a uniformly declining pipeline. In the WELAC survey, women accounted for 41.3% of Mexican undergraduate economics graduates in 2019, close to the regional average of 42.2%. Their share among Mexican master's graduates was lower, at 31.3%, compared with 38.6% across the region. Among the Mexican institutions ranked in the top five by RePEc, the corresponding shares were 30.6% and 27.7% (WELAC, 2023).[12]

[11] The 2012 constitutional amendment required the state to expand the offer of upper-secondary places gradually from the 2012/13 academic year through 2021/22 (Diario Oficial de la Federación, 2012). The obligation concerned the provision of a place; it did not imply automatic enrollment or completion. SEP describes the technological baccalaureate as a bivalent program that provides both university preparation and technical training (Secretaría de Educación Pública, 2017). This is why we do not characterize professional-technical education as categorically closing access to higher education.

[12] WELAC contacted 111 institutions in ten Latin American countries and received responses from 84. Student figures refer to 2019 graduates, whereas faculty figures refer to 2022. The report sometimes presents country totals and sometimes simple averages across responding institutions; the denominator and weighting therefore matter for comparisons (WELAC, 2023). We use the institution-average values reported in its country tables for Mexico and the country-level regional averages as contextual benchmarks.

Representation among economics faculty is also below parity, but Mexico's rank profile is not monotonic. Women represented 27.9% of assistant professors, 31.7% of associate professors, and 28.7% of full professors in the institutions covered by WELAC. The regional averages were 35.0%, 33.5%, and 24.0%, respectively.

National evidence points more clearly to stagnation and to a seniority gradient within research careers. Arceo-Gómez (2025) finds that women's representation among economics students remained between approximately 38% and 42% from 2010 to 2022. Women constitute 34% of economists recognized by Mexico's National System of Researchers (SNII), but only 20% at its highest level. The same study concludes that the educational data provide little evidence of a simple leaky pipeline, while the research-career data reveal substantially lower representation at senior levels.[13]

For this paper, the relevant fact is therefore narrower. Women are already a minority among students enrolled in or completing economics, and the imbalance has shown little sustained improvement. The experiment targets a decision made before university entry and tests one potential constraint: students' beliefs about the field and about who belongs in it.

3.3 Setting

We conducted the experiment in Nuevo León through two upper-secondary systems: the public Universidad Autónoma de Nuevo León (UANL) and the private Tecnológico de Monterrey (Tec), which operates PrepaTec. Their institutional differences motivated the stratification of school assignment by university network. The partnership gave the study access to students approaching a field-of-study decision across both networks.[14] Online Appendix Figure A6 maps the original school assignment frame, distinguishing treatment and control campuses in the two networks.

The gender imbalance motivating the intervention was visible in both university systems. Descriptive data provided by the partner institutions placed women at approximately 32% of enrolled

[13] In the SNII cross-section reported by Arceo-Gómez (2025), women account for 41% of Candidates, 28% at Level I, 27% at Level II, and 20% at Level III. These differences do not by themselves identify attrition or promotion discrimination because the ranks contain different cohorts and fields of specialization. They nevertheless provide a transparent benchmark for the scarcity of women at the top of the Mexican economics profession.

[14] Institutional records supplied during project design described 55 UANL-affiliated schools serving approximately 78,397 students and eight Tec-affiliated schools serving approximately 13,673 students. These figures describe the partner-defined networks, not the experimental sample or the number of students approached for consent; institutional directories may also count schools, campuses, and satellite units differently. The randomization included 29 schools—24 in the UANL network and 5 in the Tec network—as described in Section 4.3.

economics majors at both UANL and Tec. The setting also permits us to connect survey responses with an observed educational choice. UANL and Tec supplied applicant registries after the admissions cycle, allowing us to identify applications to economics at the two partner institutions.[15]

Nuevo León therefore provides evidence from two contrasting school systems observed at the same pre-university decision point. The sample is not intended to represent all Mexican upper-secondary students. Its value lies in combining a public and a private institutional network while preserving a common intervention, a common timing, and administrative follow-up of applications.

4 The Intervention and Experimental Design

4.1 The intervention

The intervention was designed to address women's misperceptions and beliefs regarding economics and their quantitative abilities. By engaging students at a critical stage in their educational trajectory—just before university application decisions—it aimed to reshape beliefs about economics as a field and strengthen students' confidence in their own quantitative skills.

The intervention consisted of a single, interactive classroom session led by female economists. During these in-person visits, role models delivered a 35–45 minute session that combined the sharing of personal narratives, the projection of informative videos, guided reflection exercises, and structured information delivery. Each session took place during the *Proyecto de Vida* course, a compulsory final-year class that focuses on students' personal development and career planning. This format ensured high participation rates and curricular relevance, while allowing discussions to unfold in a psychologically safe and familiar classroom environment.

Each session followed a standardized structure that combined informational, motivational, and reflective components:

- **Information about economics as a field:** an overview of the breadth of economics as a profession and its applications to issues such as poverty, inequality, health, and education, highlighting the social impact of the discipline.

[15]The registries do not observe applications to other universities, and Tec's shared entry structure allows some students to defer a program choice. To address this issue, we implement an online survey to measure application to economics programs in other universities. Section 5.2 defines the institution-specific records and explains how we construct the common application outcome.

- **Labor market opportunities:** discussion of career paths available to economics graduates in public, private, and non-profit sectors, emphasizing their diversity and relevance.
- **Challenging stereotypes and gender biases:** sensitization exercises designed to confront common misconceptions about economics being overly technical or male-dominated, and to promote the view of economics as an inclusive and impactful field.
- **Role model narratives:** personal testimonies from the visiting economists about their motivations to study economics, the challenges they faced as women in the field, and the strategies they used to succeed.
- **Future perspectives:** a short discussion on how new technologies and data-driven approaches are expanding the scope of economics, underscoring its relevance in the era of artificial intelligence and digital transformation.

All materials used during the intervention—including slides, videos, activity guides, and communication templates—were developed collaboratively by an interdisciplinary team of economists, educators, and psychologists from both universities. This team was also responsible for training the role models, ensuring consistent delivery across schools, and maintaining the fidelity of the intervention to its pedagogical and behavioral objectives. In addition, role models received preparatory training focused on communication, gender-sensitive facilitation, and classroom engagement strategies. Online Appendix Figure A1 reproduces representative slides from the implemented presentation.[16]

Role models. We recruited 14 role models for the intervention. Their ages ranged from 23 to 49 years, with a mean of 35.2. Eleven described economics as their primary field of training, two public policy, and one finance. Six held doctorates, four master's degrees, and four bachelor's degrees. They reported an average of 12 years of professional experience, ranging from 3 to 27 years, and nine had prior experience with mentoring or public outreach. Their positions spanned

[16]Online Appendix Figures A2 and A3 reproduce selected frames from the institution-specific versions of the first two videos, which introduce economics and address gendered beliefs about studying the discipline. Figure A4 documents the third video, in which women economists discuss their educational choices and professional experiences. The implementation protocol combines the three videos with the visiting role model's personal reflections and places the final video before the motivational closing. The appendix preserves the original Spanish-language images and provides links to the complete videos.

universities, government, financial services, data analytics, and private-sector strategy and pricing. This diversity was deliberate: the intervention sought to show that economics opens several professional paths rather than presenting a single model of an economist. The role models participated voluntarily and did not receive an honorarium. All received common materials and a structured delivery protocol, while the central personal narrative drew on their own career and experience.

4.2 The control condition: an active placebo

Spillover and control classrooms received a placebo session, *Elegir una carrera: una decisión sobre tu vida*, designed to provide comparable classroom contact and career-choice salience without economics or gender content. Both scripts combine a presenter introduction, discussion of the uncertainty surrounding career choice, a personal narrative, videos, a motivational closing, and time for questions. The treatment protocol includes three economics-related videos, whereas the placebo includes two videos about choosing a career. Both sessions therefore combine an outside visitor, a personal narrative, and audiovisual material delivered during a regular class hour.

The placebo contains no economics or gender content. Its script instructs the presenter to speak about the process of choosing rather than about any specific field, and explicitly to *avoid promoting a particular discipline*; the treatment script, by contrast, is organized around economics as a career and includes the role model's account of the challenges she faced as a woman in a historically male-dominated field.

The presenters also differ between arms. Treatment sessions were delivered by the female professionals recruited as role models. Placebo sessions were delivered by undergraduate research assistants drawn from the Tecnológico de Monterrey's scholarship programme (*Programa de Becarios*), who implemented the fieldwork and who were studying a range of disciplines rather than economics. In the placebo session, the placebo presenter tells the story of her or his own path into whatever field that happens to be, thereby preserving the session as a generic careers talk.

Thus, the treatment–control contrast holds constant the visit itself, the general format, the use of a personal narrative and videos, and the salience of the career decision, while varying three things jointly: the subject of the session, the presenter's field, and the presenter's career stage.[17]

[17]Notice that while the treatment presenter is an established woman working in economics or a closely related field, the control presenter is a university student, older than the average participants and typically from another discipline. The resulting estimand captures whether a classroom visit by a female economist affects outcomes relative

4.3 Experimental design and randomization

We implemented a clustered randomized controlled trial to evaluate the causal impact of the intervention on the beliefs and aspirations of high-school students towards economics. Randomization covered 29 upper-secondary-school assignment units spanning 30 campuses in the Monterrey metropolitan area—24 units affiliated with UANL and 5 Prepa Tec campuses—containing 168 classroom groups enrolled in the *Proyecto de Vida* course. One Prepa Tec campus withdrew after random assignment and three further classrooms contributed no baseline responses under their assigned identifiers (Section 4.4), so the study sample comprises 28 assignment units spanning 29 campuses (24 UANL, 4 Tec) and 159 surveyed classrooms.

Schools were the primary units of randomization. We used administrative rosters available before fieldwork to form blocks based on the share of female students and the share of classroom groups taught by women. Within each block, we randomly assigned schools to treatment and control in approximately equal numbers; when a block contained an odd number of schools, the additional school was allocated randomly between the two conditions.[18] We generated one random allocation within each block and assessed balance in the number of classroom groups recorded for each school.[19] Student-level baseline outcomes could not enter the assignment procedure because the baseline survey was administered on the day of the visit. Section 5.5 evaluates balance using the preassignment school characteristics and the observed student samples.

After assigning schools, we randomly selected up to six groups enrolled in the compulsory *Proyecto de Vida* course within each school. In treated schools, we then randomly assigned half of the selected groups to the role model session and half to the spillover condition, with one additional group assigned randomly when the number selected was odd. This classroom-level assignment was conducted separately within each treated school, without additional covariate-balance restrictions.[20]

to a structurally similar career session delivered by a university student.

[18]The blocked procedure applied to the 24 UANL units. Crossing indicators for the two gender-composition measures generated four blocks, containing seven, five, six, and six units, respectively. Because the Tec network contained only 5 campuses, we assigned them in a single block without additional covariate-balance restrictions. The resulting allocation placed 12 UANL units and three Tec campuses in treatment, and 12 UANL units and two Tec campuses in control.

[19]For UANL, the four within-block checks of the number of recorded classroom groups yielded p-values ranging from 0.150 to 0.781. These are diagnostics of the realized allocation for this school characteristic, rather than evidence of balance in every baseline variable or a criterion applied through repeated draws.

[20]Most assignment units contributed six groups. Prepa 11 contributed three, Prepa 21 four, and EIAO–Monterrey II five. The San Pedro and Santa Catarina campuses of Prepa 23 shared one school-level assignment and together contributed six groups: one in San Pedro and five in Santa Catarina. Thus, the one-group campus did not constitute

Spillover groups in treated schools and selected groups in control schools were assigned to the active placebo session described above. The initial allocation comprised 43 treatment, 42 spillover, and 83 control classrooms, before school withdrawal and other implementation losses.

This two-level randomization structure provides two complementary sources of variation: (i) a between-school comparison for the effect of assignment to a treated school, and (ii) a within-treated-school comparison for the additional effect of direct exposure to the role model session. Throughout the paper, *treatment* denotes groups assigned to the role model session, *spillover* denotes groups in treated schools that were not assigned to that session, and *control* denotes groups in control schools. Online Appendix Figure A5 summarizes the two assignment stages and the comparisons they identify.

4.4 Implementation

Fieldwork was carried out during the spring semester of the 2025/26 academic year. Classroom sessions and the accompanying in-class data collection ran from March 17 to May 21, 2026, somewhat later than the February–April window originally planned, owing to school calendars and scheduling constraints.[21] Figure 2 summarizes the calendar timing of randomization, classroom visits—during which the baseline, intervention, and immediate survey took place—the follow-up survey, receipt of the administrative records, and the online endline.

The final sample size was reduced due to departures of the original protocol during the field implementation. One Prepa Tec campus declined to participate after random assignment and was excluded with its six classrooms, all of them control. Two additional Tec control classrooms contributed no survey data, and one Tec spillover classroom had no responses recorded under its assigned classroom identifier. The initially missed visit to a UANL treatment classroom was rescheduled, and its recovered data are included. The loss therefore fell disproportionately on the Tec control group, which retains only 39 students in the main analytic sample against 137 in spillover and 153 in treatment classrooms—a point we return to in Section 5.5.[22]

a separate randomization unit.

[21] All schools assigned to treatment received the role-model intervention before the close of the relevant application window. Some control and spillover classrooms were visited after the UANL application deadline, as implementation prioritized direct-treatment sessions. This sequencing could raise concerns about spillover exposure before measurement, although we find little evidence of spillover effects in the data (Section 7.4).

[22] Some sessions did not take place as initially scheduled: one where the teacher had not been notified and the class was occupied with another project, one where fewer than ten students agreed to participate and the session was

Five classrooms received the intervention but were initially administered an old version of the measurement instrument. A later Qualtrics recovery collected the short-term instrument from some of those students, but their baseline could not be recovered. Taking advantage of the random selection of classrooms in the experiment, the research team therefore selected five replacement classrooms at random and administered the correct baseline and short-term instruments through oTree.[23]

Students could decline at any point, as requested by the IRB protocol. Therefore, a missing baseline survey may reflect ordinary or strategic absence, a decision not to participate after arriving, or an unreturned parental consent form. We observe completed surveys but cannot distinguish these reasons for non-participation. Students received equivalent invitations across experimental arms, without disclosure of the assigned session before the participation decision. The baseline survey was administered before the assigned activity began.[24]

Across 159 classrooms with baseline responses and a usable enrollment count, 3,763 of 5,635 rostered students completed a baseline survey (66.8%). Completion was 64.9% in direct-treatment, 68.3% in spillover, and 67.0% in control classrooms. The assignment roster also reports the number of women enrolled in each classroom. On that basis, 2,009 of 3,064 rostered women (65.6%) completed the baseline. Online Appendix Table A3 reports the overall and women's counts by arm.[25] These ratios measure survey completion relative to enrollment, not consent, attendance, or receipt of the intervention.

---

cancelled, one where the course had already ended because the school finished its semester early, and one where the role model was called away and the visit was rescheduled.

[23]There were other minor deviations in the field that were addressed during the implementation phase. In one treated school a session assigned to a spillover classroom was mistakenly delivered as a treatment session, and in the same visit two class groups turned out to have been merged, so students from a second group had to be removed mid-session. In two further classrooms students from another group attended, and in one case a group that had already been treated was revisited. We retain each classroom's *original roster assignment* in the analysis rather than reclassifying it according to the session delivered. The intention-to-treat estimates therefore refer to assignment under the implementation observed.

[24]Connectivity was also a recurring problem. In a substantial minority of classrooms some or all students completed the instruments on paper. In most of the cases this happened because the school had no working internet, the survey application would not work properly, or because students had no phone. Two classrooms could not play the intervention videos, one because the room had no internet and one because the screen failed seconds before the end; a third had audio problems.

[25]The denominators are enrollment counts from the assignment roster, not attendance measured on the day. We exclude classrooms without a completed baseline survey from this conditional coverage diagnostic. The confirmed transcription error for spillover classroom 108 has been corrected from 409 to 40 rostered students. The calculation includes its 27 baseline respondents and the valid responses from the five replacement classrooms. It excludes the 120 individual records from the five original classrooms affected by this measurement problem, 106 of which contained baseline responses.

Dispersion across classrooms is nonetheless wide. Five direct-treatment classrooms had baseline completion below 30% of rostered enrollment (the lowest at 11%), compared with one spillover and one control classroom. The teacher questionnaire also asks about enrollment, not attendance observed by the research team. For the 140 classrooms with both enrollment counts, 3,266 baseline respondents represent 66.9% of the 4,879 students on the assignment roster or 70.4% of the 4,638 students reported by teachers.

The study protocol was approved by the Research Ethics Committee under protocol P-ECSG-202511-01. Parents were notified in advance and could opt their children out of data collection; students provided assent at the start of each survey. All data collected from the 57 students whose parents exercised the opt-out were deleted from the study database. The trial was registered in the AEA Social Science Registry (AEARCTR-00017053) and a pre-analysis plan was filed before the analysis of any outcome data.

5 Data

5.1 Survey data

We designed three student survey instruments to measure outcomes across multiple time horizons: a *baseline survey* administered immediately before the assigned classroom session, a *short-term endline* administered immediately afterward, and a *follow-up survey* fielded several weeks later. Direct-treatment classrooms received the role-model session between the baseline and short-term endline; spillover and control classrooms received the active placebo. Thus, all participating classrooms received a session as part of the survey visit.

Baseline survey. The baseline captures pre-treatment information on demographic background, socioeconomic status, academic performance, and household proxies for educational resources. It also includes pre-treatment measures of the four main outcome domains described below: intentions and aspirations, beliefs about economics, identity and self-concept, and gender stereotypes and perceived barriers. Short lab-style items elicit percentile and probability beliefs about academic performance, providing measures of perceived academic standing and confidence.

Short-term endline. Administered immediately after the assigned role-model or placebo session, this survey measures exposure, recall, and immediate reactions to the classroom session,

as well as short-term changes in beliefs, gender stereotypes, identity alignment, and academic intentions. We use the label "short-term endline" consistently for this wave; it was completed in the same session as the baseline.

Follow-up survey. Fielded between April 30 and May 28, 2026, this wave measures persistence of treatment effects and recall of the visit. It includes the same core domains as the earlier waves, plus additional items on information requests, perceived barriers to studying economics, and interest in receiving program updates.

Most instruments were administered digitally in class: students completed the baseline and short-term endline in a single supervised digital session, using a survey application developed in oTree. Classroom instructors administered the follow-up survey through Qualtrics, while the laboratory team monitored it remotely. Two complementary administration modes were used where digital collection was not feasible. In classrooms without reliable internet access, and as a general backup, surveys were administered on paper and later transcribed. All records are linked across waves and modes through an anonymous self-generated identifier (based on surname letters, phone digits, and birth month), so that no student names were ever collected on the research instruments.

In addition to the three in-class waves, we fielded a light-touch *online endline* in August 2026, after the admission cycle, to capture students' realized post-graduation choices. Students received an email invitation to a results page where they reported whether they would continue to university and at which institution. We designed this instrument with the goal of covering students that may have decided to study in a different university than UANL and Tec. This complementary survey was administered between August 4 and August 23, 2026.

Survey items include 0–10 rating scales, categorical choices, 0–100 beliefs, and monetary decisions. Common items support longitudinal comparisons, while the questionnaires also contain wave-specific modules. English translations of the archived baseline, immediate post-intervention, and follow-up questionnaires appear in Online Appendix Sections D.1, D.2, and D.3, respectively. These sections reproduce the complete wording and response formats of the available printable versions. The complementary August questionnaire, including its display conditions and the subsequent invitation to laboratory studies, appears in Section D.4.

5.2 Administrative data

The administrative application measure uses records from both study institutions. From the UANL we obtained the registry of applicants (*aspirantes*) to the *Facultad de Economía* for the 2026 admissions cycle, including indicators for having taken the entrance exam and having been admitted. From the Tecnológico de Monterrey's admissions office we recovered the 2026 application records for the School of Social Sciences and Government. Records in both files are matched to study participants through the same anonymous identifier used in the surveys, reconstructed from fields available in each registry.[26]

We measure applications to economics using university records and students' survey responses. The administrative measure identifies students appearing in the UANL economics registry or selecting the economics programme in the Tec file. The combined measure also incorporates August-survey reports, extending coverage to students pursuing economics outside these two institutions and to cases not recovered in the registries.[27]

The follow-up survey provides a complementary but earlier measure. It asked students directly whether they had applied to an economics program. Because this measure is post-treatment, available only for follow-up respondents, and collected before choices were final, we use it as a complementary application outcome. The August online survey, fielded after the admissions cycle, supplies the survey component of the combined measure described below. Online Appendix Table A15 compares completed applications reported at follow-up with applications recorded at UANL and Tec.

Applications to economics programs at *other* institutions remain outside the two administrative sources. The online endline helps capture these choices and assess the coverage of the registries. The combined measure equals one when an economics choice is detected either in a registry or in the August survey. The survey component includes respondents who report economics at UANL or Tec but are not recovered in the corresponding registry, as well as respondents who report

[26] At both institutions, the administrative outcome identifies an explicit choice of economics at the time of application. At Tec, the LEC code records the program declared when applying, irrespective of subsequent admission or enrollment. Applicants selecting only the common entry track without declaring economics are not counted as economics applicants. Subsequent program choices or transfers fall outside the scope of this initial-application outcome.

[27] The PAP defines the primary outcome as applications to an undergraduate economics program and specifies UANL and Tec administrative records and institutional application logs as the planned sources. We report this administrative measure alongside the measure combining records and survey responses. The survey information extends the coverage of educational choices beyond these registries.

economics at another institution. We record whether each positive case is administrative only, survey only, or present in both sources. The combined-source outcome is defined for the full analysis population: zero denotes no detected event, including cases without a survey response and without an administrative match. We do not condition these regressions on endline response. The two sources provide complementary information. Registry linkage determines the coverage of the administrative measure, while the combined measure also depends on August-survey participation.

The endline was fielded from August 4, the day after the 2026/27 academic year began, and asked students whether they would be studying in the coming semester and what and where they would study. Its sample responds selectively—response is higher in the treated arms, though after the final round of e-mails the treatment−control gap is no longer statistically distinguishable from zero (Section 5.4).[28]

5.3 Sample and participation

Panel A of Table 1 describes the randomization frame and realized baseline sample, while Panels B and C assess preassignment school balance for the full set of randomized units. Table 2 reports classroom assignment and completed surveys by experimental arm and wave: Panel A pools both school networks, while Panels B and C report the Tec and UANL networks separately. Online Appendix Table A3 reports baseline-survey completion relative to enrollment among classrooms with survey data and a usable roster.[29]

In total, 3,763 students completed the baseline and 3,571 of them also completed the short-term endline (94.9% same-day retention); a further 83 students completed the short-term instrument only. The follow-up survey produced 2,767 responses, of which 1,963 could be linked to a baseline record, and the online endline was answered by 1,001 students matched to the study database.

The eligible analysis population comprises all 3,763 baseline respondents in the final year of their program: 2,666 in fourth semester and 1,079 in sixth semester, plus 18 students whose individual semester response is missing but whose classrooms belong to the eligible final-year frame. The

[28] As reported in the PAP, we were planning on implementing a difference in differences estimator with 2025 and 2026 data. However, we do not use the 2025 cohort for causal comparisons because informed consent was introduced only with the 2026 experiment, making participation across cohorts non-comparable. The PAP correspondence in the supplementary document maps the planned analyses to their current locations and distinguishes implementation changes from additional analyses.

[29] Supplementary Figure S1 shows the corresponding classroom-level distribution. These completion ratios describe coverage relative to the roster rather than treatment receipt.

primary administrative sample contains 2,009 women; pooled analyses include all 3,763 students when gender is not required. Each survey model uses the respondents with its outcome and required baseline measures, rather than imposing response at every wave. The immediate and follow-up beliefs models therefore contain 3,350 and 1,850 students, respectively.

Among baseline respondents, 3,571 also completed the short-term endline: 877 in 42 treatment classrooms, 885 in 41 spillover classrooms, and 1,809 in 75 control classrooms.[30] Table 2 distinguishes the randomized classroom roster from realized respondents. Online Appendix Table A2 reconciles the analysis populations and reports outcome-specific power benchmarks.

Students in the sample are on average 16.7 years old and 54% are female. All are in the final year of upper-secondary education, but the programs differ in length: 30% are enrolled in three-year (six-semester) programs—the Prepa Tec campuses and the UANL technical and bilingual schools—while the rest attend two-year (four-semester) programs at the general *preparatorias*. Consistent with the socioeconomic diversity of the UANL network, 13% work for pay during the school week, and fewer than half have a university-educated parent (40% of fathers, 38% of mothers). Exposure to the profession is limited: while 35% report a parent working in an economics- or business-related occupation, only 29% say they know an economist personally.

Baseline attitudes display the pattern that motivates the intervention. Students hold a distinctly prosocial view of the discipline—the belief that economics improves people's lives averages 7.5 on the 0–10 scale, and perceived family support for studying economics averages 8.6—yet their personal engagement with the field is weak: liking of economics averages 5.3 (below the 5.7 for mathematics), self-assessed ability in economics is 4.9 (against 6.1 in mathematics), and the self-reported likelihood of applying to an economics program is just 3.6. In other words, students see economics as socially valuable but not as a field for themselves—precisely the wedge between the perceived value of the discipline and perceived personal fit that the role-model sessions were designed to close. Figure 3 reports the corresponding female–male baseline differences for the component measures and domain indices, grouped into beliefs about economics, academic orientation and self-concept, intentions and aspirations, and gender stereotypes. Online Appendix Table A4 reports the corresponding

[30] One of the 159 classrooms with baseline responses contributes no student to this paired sample: a treatment classroom in which no student completed both instruments. Separately, a spillover classroom on the assignment roster was never visited due to coordination issues and contributes no survey data. The paired count describes same-day survey completion; it does not define eligibility for administrative or follow-up analyses. Gender-interaction models require observed gender and consequently omit the 65 eligible baseline respondents without an observed classification.

descriptive means and gender differences.

5.4 Attrition

Online Appendix Table A25 tests participation after baseline within the eligible final-year population. Same-day response is 95.0 percent in control schools, and the treated-school difference is 0.6 p.p. (S.E. = 1.3), which is not statistically significant. Follow-up response is 48.3 percent in control schools and 13.7 p.p. higher under treated-school assignment (S.E. = 7.0), a statistically significant difference at 10 percent. The online-endline difference is 3.1 p.p. (S.E. = 2.5) from a control mean of 22.8 percent and is not statistically significant. The follow-up imbalance makes selection especially relevant to the persistence results. Section 10.1 therefore examines inverse-probability weighting and Lee bounds. Both application measures are defined for all eligible baseline respondents; the combined measure also uses August-survey reports. Assignment preceded classroom consent, so observed participation also reflects the consent and implementation protocol.

5.5 Balance

Table 1 assesses comparability at the level at which schools were assigned. Panel B reconstructs school-unit characteristics from the preassignment *Proyecto de Vida* classroom rosters for all 29 randomized units. Panel C adds information assembled in 2025 for the 24 UANL units, including total enrollment and the number and gender composition of *Proyecto de Vida* instructors. Unlike a comparison among 2026 survey respondents, neither panel conditions on consent, attendance, or survey completion.

We report unadjusted control and treated-school means together with standardized differences that respect the blocked assignment. For each characteristic, we estimate the treated-school difference conditional on the original randomization-block fixed effects and divide that coefficient by the pooled within-arm standard deviation. We assess balance from the magnitude of this statistic rather than from hypothesis tests, whose values depend on sample size and are particularly uninformative with 29 school-level assignment units.

The preassignment comparisons indicate broad, but not exact, balance. In the full randomization frame, the absolute standardized differences for the number of eligible classrooms, the number of eligible students, mean class size, the female student share, and the share of classes taught by

women range from 0.118 to 0.208. The exception is scheduling: treated schools have a larger share of eligible classes in the morning shift, with an adjusted standardized difference of 0.380. The additional 2025 UANL measures are similarly close for enrollment, the number of *Proyecto de Vida* groups, and the female instructor share, but treated units have more *Proyecto de Vida* instructors (0.317 standard deviations). We therefore describe the assignment as reasonably balanced while retaining the original randomization-block fixed effects in all school-level specifications.

We also examined structural characteristics reported by school directors. Those responses were collected before exposure during the school visit but after assignment. They are therefore useful as a supplementary check on predetermined school structure, but not as the primary test of the randomization. Most of their absolute standardized differences are below 0.25. One exception is the share of degree-holding teachers (0.319).

5.6 Outcomes and index construction

We examine applications to an undergraduate economics program during the 2026 admissions cycle using administrative records and a measure combining those records with August-survey reports (Section 5.2). The administrative indicator equals zero when neither registry identifies an economics application; the combined indicator also counts economics choices reported in the August survey.

Secondary outcomes capture four pre-specified domains through which role models may affect applications: intentions and aspirations, beliefs about economics, identity and self-concept, and gender stereotypes and perceived barriers (Breda et al., 2023). Their measurement depends on the survey wave. The immediate survey supports multi-item summaries of beliefs and intentions, but measures identity through perceived fit and gender stereotypes through a single endorsement item. The follow-up supports all four domain summaries. Recall of a female economist and reported barriers are analyzed separately from the stereotype index.

We orient items toward greater alignment with the target construct and combine them into inverse-covariance-weighted indices using baseline control-group moments and weights (Anderson, 2008). Effects are expressed in standard deviations with respect to the control group. Benjamini–Hochberg (Benjamini and Hochberg, 1995) q-values adjust the four domain tests within each wave and estimand. Online Appendix Section B details the item composition, weighting, and treatment

of missing responses. Tables A5–A7 report coverage and reliability by wave.

6 Empirical Strategy

We estimate intention-to-treat effects for the 2026 experimental cohort, retaining students in their assigned treatment groups regardless of reported exposure. The school-level ITT compares all students in treated schools, including those in spillover classrooms, with students in control schools. Within treated schools, classroom assignment allows us to separate direct exposure from potential spillovers. We denote classrooms assigned to the role-model session by T, classrooms assigned to the placebo in treated schools by S, and classrooms assigned to the placebo in control schools by C. The T–C contrast measures the total direct effect, T–S the incremental effect of the role-model session within a treated school, and S–C the spillover contrast.[31]

The main specifications use unweighted student-level regressions, with each observed student contributing one observation. This weighting reflects our focus on students' beliefs and application decisions in the analysis sample.

6.1 Primary outcome: populations and endpoints

Our primary outcome is whether a student applied to an undergraduate economics program at any university. We combine UANL and Tec records with responses to the August survey, which asked students what and where they expected to study. The indicator equals one if either source identifies an economics choice and zero otherwise. The survey thus extends coverage to universities outside the two registries. Our primary sample consists of women observed at baseline in their final semester of high school. We also report estimates for all students observed at baseline in their final semester. Both samples include students in the fourth semester of two-year programs and the sixth semester of three-year programs. For comparison, we report the administrative measure for UANL Economics and Tec's LEC program combined and then for each institution separately. A broader administrative measure combines applications to UANL Economics with applications to Tec's ECSG gateway, which covers several disciplines.

[31] The pre-registered comparison with the 2025 cohort cannot be implemented under a common selection rule. Consent was requested only in 2026, after the field team arrived in the classroom. Comparing the two cohorts would therefore combine the intervention with differences in consent-based selection. We use the randomized 2026 cohort for all treatment-effect estimates.

We estimate linear probability models with randomization-block fixed effects. Application indicators are multiplied by 100, so coefficients and standard errors are in p.p. and control means are percentages. The main specification contains no additional covariates. Adjusted specifications add baseline age, GPA, application propensity, interest in economics, and mathematics self-assessment, using mean imputation and missing-value indicators. For outcomes incorporating August-survey reports, we retain the full eligible sample without adding a response indicator. Survey response is itself a post-assignment outcome; conditioning on it would alter the assignment comparison without resolving incomplete detection of educational choices.

6.2 Survey outcomes: ANCOVA

For the survey outcomes, we estimate separate effects at the immediate and follow-up waves, controlling for the corresponding baseline measure to improve precision (McKenzie, 2012). The school-level ITT specification is

$$Y_{sgi,1} = \alpha + \gamma_{b(s)} + \beta \, \text{TreatSchool}_s + \rho Y_{sgi,0} + X'_{sgi}\theta + \varepsilon_{sgi}, \tag{3}$$

where $Y_{sgi,1}$ is the post-treatment outcome for student i in classroom g at school s, and $Y_{sgi,0}$ is its wave-matched baseline value. TreatSchool_s indicates assignment to a treated school, and $\gamma_{b(s)}$ denotes the fixed effect for block $b(s)$, where $b(s)$ denotes the randomization block containing school s. The main specification includes the baseline outcome and block fixed effects and omits X_{sgi}. Section 10.2 examines the sensitivity to additional baseline controls and alternative index constructions.

To estimate the two cross-school contrasts, we distinguish direct-treatment and spillover classrooms:

$$Y_{sgi,1} = \alpha + \gamma_{b(s)} + \beta_1 \, \text{TreatGroup}_{sg} + \beta_2 \, \text{SpillGroup}_{sg} + \rho Y_{sgi,0} + X'_{sgi}\theta + \varepsilon_{sgi}. \tag{4}$$

TreatGroup_{sg} identifies classrooms assigned to the role-model session and SpillGroup_{sg} identifies spillover classrooms in treated schools. Both indicators equal zero in control schools. Thus, β_1 estimates T–C and β_2 estimates S–C. We estimate T–S separately within treated schools, replacing

the block fixed effects with school fixed effects. This comparison uses variation between classrooms of the same school and need not equal the difference between the two cross-school coefficients.

6.3 Inference

We cluster standard errors at the assignment level relevant to each contrast: schools for the school ITT, T–C, and S–C; classrooms for T–S, which includes school fixed effects. We report estimated coefficients with clustered standard errors and assess statistical significance using two-sided p-values. For multiplicity-adjusted families, the corresponding criterion is a q-value below 0.10. Unless stated otherwise, statistical significance refers to conventional clustered inference; we identify wild-bootstrap, randomization-inference, and adjusted tests explicitly. Benjamini–Hochberg adjustments apply to the four survey domains within each wave and estimand. Item-level and heterogeneity analyses use their specified families; tables distinguish adjustments of conventional tests from adjustments of wild-bootstrap tests.

Because the number of school clusters is small, we supplement conventional inference with null-imposed wild cluster bootstrap-t tests and randomization inference. The bootstrap uses Rademacher weights, symmetric two-sided tests, and 5,000 replications. Randomization inference uses 9,999 reassignments and add-one p-values. School reassignments preserve the realized treatment counts within the five randomization blocks, and the classroom stage assigns a balanced number of direct-treatment classrooms within the re-drawn treated schools. For T–S, treated schools remain fixed and classroom labels are permuted while preserving their observed treatment counts.[32] We assess the principal findings against these procedures and use leave-one-school-out estimates to diagnose influence.[33]

[32] Randomization draws use the full original frame of 29 school assignment units and 168 classrooms, including units without usable observations. Outcomes, baseline covariates, and the realized analysis sample remain fixed; units without data contribute only to the assignment draws. The withdrawn Tec school's assignment can therefore vary across draws. These tests do not correct selection into the observed sample. For T–S, the within-school permutation test supplements conventional inference, which assumes independence across classrooms conditional on school fixed effects.

[33] Online Appendix Section C.2 and Table A2 report outcome-specific power benchmarks for the realized design. The approximate minimum detectable effect for women's combined application measure is 0.89 p.p., compared with 0.76 p.p. for the administrative measure. The corresponding benchmarks for beliefs are 0.165 standard deviations immediately after the intervention and 0.144 at follow-up. These descriptive power benchmarks retain their original two-sided 5 percent test and 80 percent power.

7 Results

We first examine whether the intervention changed beliefs and related survey outcomes among final-year students, then assess persistence and applications to economics. We conclude by separating direct effects from spillovers to other classrooms in treated schools. All principal estimates pool eligible students from four- and six-semester programs.

7.1 Effects on beliefs and identity

The clearest immediate effect is an improvement in beliefs about economics. Panel A of Figure 4 reports the overall school ITT alongside the effects by gender. The beliefs index increases by 0.301 standard deviations with respect to the control group (S.E. = 0.059). The coefficient is statistically significant, survives the four-domain adjustment (q-value < 0.001), and remains statistically significant under the school-level wild bootstrap and randomization inference (p-value < 0.001 in both cases). The intervention therefore changes students' reported understanding of the field even when inference accounts for the small number of schools.

The immediate identity measure also increases, by 0.097 standard deviations (S.E. = 0.025). The coefficient is statistically significant after the four-domain adjustment (q-value = 0.001) and passes the wild-bootstrap test (p-value = 0.004) and randomization inference (p-value = 0.004). The estimated coefficients for intentions and aspirations and rejection of gender stereotypes are -0.038 (S.E. = 0.065) and 0.052 (S.E. = 0.037), respectively; neither is statistically significant. Thus, the immediate response combines improved beliefs about economics with a smaller improvement in perceived fit.[34]

Panel B of Table 3 shows positive immediate belief effects for both men and women: 0.268 (S.E. = 0.071) and 0.342 (S.E. = 0.067) standard deviations, respectively. Both coefficients are statistically significant. The formal women-minus-men difference is 0.074 (S.E. = 0.069) and is not statistically significant (conventional-test q-value = 0.446), so the larger estimate for women does not establish a gender difference. Immediate perceived fit shows a more differentiated response: the estimate is 0.017 for men and 0.157 for women, with a statistically significant difference of 0.140

[34] Online Appendix Table A38 reports the corresponding tests for every domain and assignment contrast. Supplementary Figure S11 contrasts the significant belief effect with the nonsignificant application estimate: the immediate beliefs estimate lies in the tail of its randomization distribution, while the women's administrative application estimate does not. The two panels use different outcome units.

(S.E. = 0.051). This interaction survives the conventional four-domain adjustment (q-value = 0.044) and the wild-bootstrap adjustment at the 10 percent level (q-value = 0.070), supporting a larger identity response among women.

The item-level results clarify what changed. Students become more likely to reject the view that economics is only about money and finance and to report understanding economists' work. The respective coefficients are 0.225 (S.E. = 0.050) and 0.178 (S.E. = 0.077) standard deviations. Both are statistically significant, with beliefs-family q-values below 0.001 and 0.044 (Online Appendix Table A8). The estimated increase in perceived social value is smaller and not statistically significant. Tables A9–A11 show that the immediate identity response reflects perceived fit, while application likelihood, expected family support, and rejection of the gender stereotype do not change detectably.[35]

The effects are concentrated in beliefs about economics and perceived fit. At follow-up, the positive effect on beliefs persists, while interest in English, history, and Spanish shows no detectable response (Online Appendix Table A39, Panel A). Together with the active-placebo comparison, this pattern provides evidence against a generalized tendency to report more favorable outcomes. It is consistent with updating specific to the intervention's content, but does not distinguish such updating from content-specific experimenter demand. Online Appendix Section C.1 discusses these complementary checks.

7.2 Persistence at the follow-up wave

The belief effect remains positive at follow-up. Panel B of Figure 4 reports a school-level estimate of 0.169 standard deviations (S.E. = 0.052), compared with 0.301 immediately after the intervention. The follow-up coefficient is statistically significant under conventional inference, the wild bootstrap (p-value = 0.009), randomization inference (p-value = 0.008), and the four-domain adjustment (q-value = 0.012). We find no evidence of follow-up effects on identity, intentions, or gender stereotypes. Panel A of Table 3 shows positive belief effects for both genders, with a women-minus-men difference of 0.067 standard deviations (S.E. = 0.072), which is not statistically

[35] Supplementary Table S1 separates item effects by gender and assignment contrast. The school-ITT effect on understanding economists' work is 0.209 standard deviations larger for women (S.E. = 0.083), a statistically significant difference after adjustment (wild-bootstrap q-value = 0.079). The identity-fit difference also survives its item-family adjustment. These findings identify selected gender differences, rather than a uniformly larger response across all immediate outcomes.

significant. The follow-up response is therefore concentrated in beliefs about economics.

The difference between the two index estimates is not itself a test of attenuation. Online Appendix Table A12 and Figure A7 hold respondents and item definitions fixed to compare the waves directly. In this balanced panel, the effect on rejecting the money-and-finance stereotype is 0.158 standard deviations immediately and 0.139 at follow-up; the change is −0.019 (S.E. = 0.046) and is not statistically significant. None of the seven common-item changes survives multiplicity adjustment. Several estimates decline, but the formal comparisons do not establish a common pattern of decay.[36]

Among the follow-up belief items, rejection of the money-and-finance stereotype increases by 0.116 standard deviations (S.E. = 0.049), a statistically significant coefficient that also passes the beliefs-family adjustment at the 90 percent confidence level (q-value = 0.077; Online Appendix Table A8). We find no statistically significant effects on the other belief components or the individual identity, intentions, and stereotype items (Tables A9–A11).[37]

Incomplete follow-up and differences in measurement timing across arms remain relevant. The balanced panel conditions on response at both waves and does not remove those timing differences. Section 10.1 shows that reweighting preserves the positive follow-up belief estimate, while the Lee bounds still include zero. We therefore interpret the results as evidence of a response at the observed follow-up, with remaining uncertainty about selection and longer-term persistence.

7.3 Applications to economics

Table 4 examines whether the change in beliefs translated into applications. Panel A reports women in the final year and Panel B all final-year students. Column 1 presents the effects on applications to economics using a measure that combines administrative records from UANL and Tec with students' self-reports, including applications to other universities. Column 2 uses only the administrative records from the two institutions. Columns 3 and 4 report the results for

[36] Online Appendix Table A29 compares each wave's full respondent sample with the balanced panel. Immediate item estimates differ across these samples, reinforcing the importance of fixing composition when studying persistence. Supplementary Tables S5–S7 and Figure S2 extend the comparisons by gender. No reported gender difference in the common-item effects or their changes survives the wild-bootstrap adjustment. These tests address different questions from whether a wave-specific effect is individually significant.

[37] The gender disaggregation in Supplementary Tables S2–S4 does not identify a follow-up item interaction that survives the reported wild-bootstrap adjustment. Thus, the index-level belief result is accompanied by a significant change in rejection of the money-and-finance stereotype, but no significant changes in the other belief items or adjusted gender differences.

applications to UANL and Tec, respectively. Finally, column 5 reports the effects on applications to other universities, measured through students' self-reports.

For the combined measure in column 1, the estimated effect of treated-school assignment is 0.333 p.p. for women (S.E. = 0.318), relative to a control mean of 0.294 percent. For all students, the estimate is 0.490 p.p. (S.E. = 0.387), relative to a control mean of 0.630 percent. Neither effect is statistically significant. This conclusion is unchanged under wild-bootstrap and randomization inference, and when we adjust for baseline covariates.[38]

Using only administrative records from the two institutions, column 2 yields an estimated effect of 0.404 p.p. for women (S.E. = 0.271), relative to a control mean of 0.196 percent. The coefficient is not statistically significant under conventional inference, the wild bootstrap (p-value = 0.182), or randomization inference (p-value = 0.211). Adjusting for baseline covariates gives 0.432 p.p. (S.E. = 0.275), also not statistically significant. The conventional 95 percent confidence interval for the unadjusted estimate is $[-0.153, 0.961]$ p.p. The confidence interval remains compatible with economically meaningful increases in women's applications. For all students, the estimate is 0.412 p.p. (S.E. = 0.325) and is likewise not statistically significant.[39]

However, the institution-specific results reveal an increase in applications to UANL among all students. Column 3 reports an effect of 0.659 p.p. (S.E. = 0.254), relative to a control mean of 0.315 percent. This result is statistically significant under conventional inference and the wild bootstrap (p-value = 0.020), and remains significant under randomization inference (p-value = 0.039). For women, the UANL estimate is positive but not statistically significant. Online Appendix Tables A13–A14 report the institution-specific results for both populations.[40]

Column 4 shows no statistically significant effect on applications to the Tec economics program. The estimated effect is 0.031 p.p. for women (S.E. = 0.036) and −0.193 p.p. for all students (S.E.

[38]Online Appendix Table A16 identifies two survey-only applications alongside eight administrative cases among women, and nine survey-only applications alongside 28 administrative cases in the pooled sample. In a separate comparison, Table A15 shows that applications reported in the earlier follow-up survey overlap only partly with those recorded by the universities. High agreement dominated by shared negative classifications cannot establish that either source captures all applications.

[39]Supplementary Table S57 holds coefficients fixed and changes the clustering level. The classroom-clustered p-value for women's administrative applications is 0.188, compared with 0.148 under the assignment-aligned school clustering. Online Appendix Figure A8 shows that omitting one school leaves this estimate between 0.261 and 0.540 p.p., while the immediate belief effect remains between 0.265 and 0.343 standard deviations. The belief result remains significant throughout; the application estimates remain positive, but their statistical significance varies with the omitted school.

[40]This UANL outcome is measured among students from both school networks; it is not an analysis restricted to UANL high schools.

= 0.164). Neither conclusion changes when we adjust for baseline covariates.

Column 5 reports small positive estimates for self-reported applications to other universities: 0.031 p.p. for women (S.E. = 0.036) and 0.096 p.p. for all students (S.E. = 0.058). Neither is statistically significant in the unadjusted specification. With baseline covariates, the all-student estimate is 0.112 p.p. (S.E. = 0.066), statistically significant at the 10 percent level under conventional inference. This finding therefore depends on the covariate adjustment; the corresponding estimate for women remains statistically insignificant.

Baseline application propensity and interest in economics positively predict administrative and combined-source applications, while GPA, age, and mathematics self-assessment have no statistically significant associations conditional on those measures.[41]

Table 5 follows the transition from applications to admission and enrollment. Panel A shows no statistically detectable gender difference in administrative, combined-source, or gateway applications, or in economics-degree enrollment. Panel B likewise shows no statistically significant gender differences for the UANL stages. In Panel C, the Tec ECSG enrollment effect is 0.761 p.p. larger for women (S.E. = 0.480), but the difference is not statistically significant (q-value = 0.873). Online Appendix Table A17 reinforces the absence of an adjusted gender difference in applications.[42]

7.4 Within-school spillovers

Tables 6–8 distinguish direct and spillover effects by gender for immediate outcomes, follow-up outcomes, and applications. Figures 5 and 6 place these effects alongside the pooled estimates. The decomposition shows that the belief response is concentrated in directly treated classrooms.

For immediate beliefs, the T–C coefficient is 0.579 standard deviations (S.E. = 0.076) and T–S is 0.540 (S.E. = 0.062). Both are statistically significant, with p-value < 0.001 under conventional

[41] Supplementary Table S11 reports these associations for women and all eligible students under both application measures. The coefficients describe conditional predictiveness rather than causal effects of the baseline characteristics.

[42] Supplementary Figure S4 displays the subgroup estimates under the combined, administrative, and gateway definitions. Subgroup coefficients come from pooled gender-interaction models with common block effects and need not equal regressions restricted to women. Supplementary Tables S8–S10 extend the administrative stages across assignment contrasts. School-assignment effects on economics-degree and gateway enrollment are not statistically significant for women or all students. For women, however, direct-treatment-versus-spillover effects are positive and statistically significant under both definitions. Pooled UANL acceptance increases by 0.346 p.p. under school assignment (S.E. = 0.183), a secondary result that is statistically significant at 10 percent under conventional inference but does not establish a general progression to enrollment. UANL acceptance is not matriculation, and Tec ECSG enrollment is not enrollment in an economics degree alone. The combined and outside-registry application measures in Table S10 are estimated without conditioning on August-survey response.

inference, the wild bootstrap, and randomization inference. S–C is 0.037 (S.E. = 0.054) and is not statistically significant (Online Appendix Tables A18 and A38). Table 6 shows larger direct effects for women: the T–C gender-difference coefficient is 0.266 standard deviations (S.E. = 0.087), and the T–S difference is 0.355 (S.E. = 0.091). Both are statistically significant after adjustment (wild-bootstrap q-value = 0.010 and q-value < 0.001, respectively). The T–C identity difference also survives adjustment (q-value = 0.027). Thus, gender differences are clearer for direct exposure than for the school ITT, which averages direct-treatment and spillover classrooms.

At follow-up, T–C is 0.305 standard deviations (S.E. = 0.061) and T–S is 0.301 (S.E. = 0.049). Both coefficients are statistically significant under conventional inference, the wild bootstrap, and randomization inference (p-value < 0.001 throughout). S–C is 0.050 (S.E. = 0.061) and is not statistically significant (Online Appendix Table A19). Table 7 again shows larger direct belief point estimates for women. However, the gender differences no longer pass the multiplicity-adjusted wild-bootstrap tests: q-value = 0.112 for T–C and q-value = 0.117 for T–S. The later response is therefore concentrated in direct-treatment classrooms, with no statistically significant gender difference after adjustment.

For women's administrative applications, T–C is 0.440 p.p. (S.E. = 0.304) and is not statistically significant. No application interaction in Table 8 survives the reported wild-bootstrap adjustment. Online Appendix Table A20 likewise shows no statistically significant administrative or combined-source application effects for women. The all-student administrative T–C coefficient is 0.817 p.p. (S.E. = 0.484), but statistical significance at 10 percent holds only under randomization inference and differs across conventional (p-value = 0.103), wild-bootstrap (p-value = 0.130), and randomization tests (p-value = 0.049). For the broader measure that includes the Tec gateway, the corresponding T–C effect is 0.847 p.p. (S.E. = 0.512). It also reaches the 10 percent threshold only under randomization inference (p-value = 0.084, compared with 0.110 for the conventional test and 0.144 for the wild bootstrap). Women's positive T–C estimate for the broad gateway also remains sensitive to the inference procedure.[43] The clearest secondary administrative response is again applications to UANL among all students: T–C is 1.133 p.p. (S.E. = 0.428) and is statistically significant, with wild-bootstrap p-value = 0.011, and randomization-inference p-value = 0.003

[43]Supplementary Figure S3 displays the administrative and gateway application estimates for women and all eligible students across the assignment contrasts. The figure complements the formal inference comparisons in Online Appendix Table A38.

(Table A38). Thus, direct exposure increases applications to UANL in the combined sample, while we find no statistically significant effect on women's administrative application measure.[44]

8 Mechanisms

The intervention combines information about economics with exposure to female role models. Our design identifies the effect of this bundle, but does not separately randomize its components. We assess its possible channels through baseline characteristics that may shape students' responses, complementary measures of perceived constraints, and an exploratory mediation exercise linking immediate responses to women's subsequent economics applications.

8.1 Baseline beliefs, interest, and mathematics self-assessment

Table 9 examines three pre-specified characteristics relevant to the proposed channels: baseline beliefs about economics, interest in the field, and mathematics self-assessment. If the intervention primarily fills information gaps, students with less favorable initial beliefs or less interest may respond more strongly. Mathematics self-assessment captures a related perceived constraint on pursuing economics. Each panel reports subgroup effects and the formal high-minus-low interaction.

The immediate belief effect is positive in every subgroup. Panel A reports effects of 0.376 standard deviations below the baseline-beliefs split and 0.227 above it, with a difference of -0.149 (S.E. $= 0.073$). This interaction is statistically significant at 10 percent under conventional inference but not after adjustment (wild-bootstrap q-value $= 0.396$). Panel B shows a similar pattern by initial interest: the effect is 0.445 for the low-interest group and 0.221 for the high-interest group. The difference is -0.223 (S.E. $= 0.078$) and is statistically significant under conventional inference, although its wild-bootstrap q-value is 0.110. Panel C shows little differentiation by mathematics self-assessment: the interaction coefficient is 0.021 (S.E. $= 0.076$) and is not statistically significant.

[44] Online Appendix Table A1 and Supplementary Table S13 document stronger recall after direct assignment, but no statistically significant effects on discussion with family, friends, or peers. Supplementary Table S12 describes reported attendance and reception of the visit. Among respondents in directly assigned classrooms, mean enjoyment is 8.17 for men and 8.56 for women on a 0–10 scale; agreement that perceptions of economists' work changed averages 7.24 in both groups. These are unadjusted descriptions of respondents' experiences. They do not measure treatment effects or establish that the visit improved perceptions. Affirmative attendance reports also occur in control and spillover classrooms, and response coverage differs across arms, limiting their interpretation as verified exposure. At follow-up, Tables S14 and S15 extend the exposure and recall comparisons by gender, with no statistically significant T–C gender differences in reported activity or exact recall.

The estimates are compatible with a larger immediate response where initial information or interest was lower, but do not establish this pattern after multiplicity adjustment.

Follow-up beliefs are positive in all six subgroups, with no statistically detectable interactions. We also find no statistically significant differences in women's administrative application effects across the three moderators. Extending the analysis to the complete outcome families in Online Appendix Tables A21–A23 yields no interaction that survives the multiplicity-adjusted wild bootstrap. The broader coverage therefore does not identify a particular baseline constraint through which the intervention changes applications.

The exploratory barrier and information-request measures provide little evidence of a general relaxation of perceived constraints. Neither the distance-barrier response nor requests for program information changes detectably under school assignment.[45]

At follow-up, we find no statistically detectable pooled effects on reported academic expectations or performance on the brief quantitative quiz. Gender comparisons nevertheless indicate a more positive response in women's reported expectations under treated-school assignment.[46] We also find no statistically detectable effect on self-assessed mathematical ability under any of the four assignment contrasts.[47]

[45]Supplementary Table S16 reports a distance-barrier estimate of −1.272 p.p. (S.E. = 2.054) and a brochure-request coefficient of 2.292 p.p. (S.E. = 2.264); neither is statistically significant. Selecting inability to imagine an economics career as a barrier increases by 5.706 p.p., but narrowly fails the within-family 10 percent adjustment (q-value = 0.101). Tables S17 and S18 do not identify gender differences that pass the wild-bootstrap adjustment in these families. The follow-up instrument permitted multiple selections and expanded the PAP response options. These exploratory outcomes measure reported constraints and demand for information, not identified mediators.

[46]Supplementary Table S19 reports a quiz-score effect of 0.256 correct answers (S.E. = 0.183), on a 0–6 scale and against a control mean of 3.646; this coefficient is not statistically significant. Neither the expected classroom percentile nor the subjective probability of exceeding the classroom median passes the original multiplicity adjustment. In Table S20, the treated-school assignment effect on expected percentile is 7.812 points larger for women (S.E. = 2.012); the corresponding difference in above-median probability is 8.744 points (S.E. = 2.664). Both differences are statistically significant after adjustment (wild-bootstrap q-value = 0.011 and q-value = 0.022, respectively). These q-values apply the Benjamini–Hochberg adjustment to wild-bootstrap tests across the original nine outcomes. Table S21 shows no statistically significant gender differences in observed quiz performance under the same assignment contrast. The expectation questions concern a hypothetical classroom exam, whereas observed ranks use sample-wide scores on a separate six-item quiz. These different benchmarks do not allow us to assess the accuracy of students' expectations.

[47]Supplementary Table S22 reports the pooled estimates, and Table S23 reports the corresponding gender comparisons. The women-minus-men differences are 0.443 points under treated-school assignment (S.E. = 0.219) and 0.504 points for direct treatment relative to control (S.E. = 0.248), on the 0–10 response scale. Both coefficients are statistically significant at 10 percent under conventional inference. These differences reflect more negative estimates for men, rather than clear gains for women. Neither contrast passes the wild-bootstrap adjustment (q-value = 0.321 and q-value = 0.437, respectively), and the other gender contrasts are not statistically significant. The item measures perceived competence rather than performance on a mathematics test.

8.2 Exploratory mediation analysis

We next examine whether immediate changes in beliefs, perceived fit, intentions, and stereotypes are associated with women's subsequent applications to economics. We consider both the administrative and broad-gateway application measures and estimate a separate model for each mediator. For each treatment contrast, we multiply the estimated effect of assignment on the mediator by the estimated association between the mediator and applications, conditional on assignment and baseline controls. Both equations include randomization-block fixed effects.[48] Because the mediators enter separate models, these products cannot be added as independent contributions. A causal interpretation requires sequential ignorability of the mediator–outcome relation, which treatment randomization alone does not establish.

Only the coefficient products involving immediate perceived fit are statistically significant after adjustment for multiple testing. For assignment to a treated school, the estimated product is 0.077 p.p. (S.E. = 0.030, q-value = 0.036) for women's administrative applications and 0.080 p.p. (S.E. = 0.036, q-value = 0.0997) for the broad-gateway measure. Both estimates are also positive and statistically significant at the 10 percent level when we compare directly treated classrooms with control classrooms. For beliefs about economics, the school-assignment product for administrative applications is similar in magnitude at 0.076 p.p. (S.E. = 0.063), but is not statistically significant (q-value = 0.455). The estimates for intentions and stereotypes are close to zero. We interpret the finding for perceived fit as an association with subsequent applications. The total effect of school assignment on women's administrative applications is not statistically significant.

9 Heterogeneity

Tables 10 and 11 summarize heterogeneity by student characteristics and classroom context. The first covers gender, parental education at the student and school levels, household assets, and baseline GPA. The second examines classroom composition, classroom GPA, baseline stereotypes, and teacher gender.

Panel A of Table 10 reports positive immediate belief effects for both genders: 0.268 standard

[48] Supplementary Tables S24–S27 report the estimates. We use the delta method with the joint school-clustered covariance matrix for the standard error of each product. The q-values adjust for testing the four mediators separately for each outcome definition and treatment contrast.

deviations for men and 0.342 for women. The women-minus-men interaction is 0.074 (S.E. = 0.069) and is not statistically significant (wild-bootstrap q-value = 0.477). Follow-up beliefs are also positive for both groups, with no statistically significant gender difference.[49]

Panels B–E of Table 10 extend the comparison to parental university education, school-level parental education, household assets, and baseline GPA. Immediate belief effects are positive throughout. For example, Panel B reports 0.311 standard deviations without a university-educated parent and 0.284 with one, a difference of −0.027 (S.E. = 0.076), which is not statistically significant. Panel E shows a 0.118-standard-deviation larger immediate belief response at high GPA (S.E. = 0.067). This interaction is statistically significant at 10 percent under conventional inference but not after adjustment (wild-bootstrap q-value = 0.288). Follow-up beliefs and women's administrative applications show no statistically significant differences across these characteristics.

The complete survey families lead to a similar conclusion. The main exception is the immediate gender difference in perceived fit discussed in Section 7.1, which passes the wild-bootstrap adjustment at 10 percent (Online Appendix Table A24). The additional outcomes do not establish that the intervention systematically benefits students at one end of the GPA distribution.[50]

Table 11 also yields limited adjusted evidence of heterogeneity. Panel A shows a smaller immediate belief effect in classrooms with a higher female share: the difference is −0.228 standard deviations (S.E. = 0.094) and is statistically significant under conventional inference, but the wild-bootstrap q-value is 0.187. Panel B shows positive belief differences by classroom GPA, while Panel C shows negative differences by baseline stereotypes; neither pattern is statistically significant.[51]

[49]Supplementary Tables S36–S37 examine whether women's relative response to the intervention varies with students' initial beliefs, interest in economics, mathematics self-assessment, or classroom teacher gender. The comparisons ask whether informational and vocational starting points and teaching context help explain gender differences in treatment responses. The follow-up belief gender difference is 0.356 standard deviations larger in the high-baseline-beliefs group (S.E. = 0.149), a statistically significant interaction at 10 percent after adjustment (q-value = 0.097); none of the other interactions survives adjustment at 10 percent. The q-values adjust conventional school-clustered p-values across four domains separately for each moderator and wave, rather than jointly across all 32 interactions. This isolated finding is suggestive and does not establish a general pattern of differential gains by gender. These exploratory comparisons identify neither the effect of changing a grouping characteristic nor the mechanism behind a treatment-effect difference.

[50]Supplementary Tables S28–S31 extend the parental-education, socioeconomic, and GPA comparisons to the complete survey families. Figure S5 shows no consistent GPA gradient across domains: for immediate intentions, the high-minus-low contrast is −0.038 standard deviations (S.E. = 0.073) and is not statistically significant (wild-bootstrap q-value = 0.800). Figure S6 likewise shows no statistically significant GPA differences in women's administrative and gateway application effects.

[51]Supplementary Figure S7 makes these directions clear without establishing a common classroom gradient. None of the application differences in Figure S8 is statistically significant. The complete classroom families in Tables S32–S34 yield no interaction that survives the wild-bootstrap adjustment.

We also examine whether the intervention's effects vary with the gender of the classroom teacher (Panel D). This exploratory comparison yields a 0.345-standard-deviation larger immediate belief effect in classrooms with female teachers (S.E. = 0.092), while the follow-up difference is close to zero. The women's administrative application contrast is negative, at −2.257 p.p. (S.E. = 0.914). Both coefficients are statistically significant under conventional inference. Neither interaction remains statistically significant at the 10 percent level after adjustment (q-value = 0.112 and q-value = 0.418, respectively). These estimates do not establish a consistent pattern of larger intervention effects in classrooms with female teachers.[52]

Exploratory heterogeneity by altruism, generalized trust, time-related attributes, and risk preferences yields no interaction that survives the original adjustment within wave and outcome-type families. These comparisons do not identify a stable subgroup pattern beyond the selected gender responses discussed above.[53]

We next examine whether baseline characteristics predict differences in treatment gains using cross-fitted LASSO and generalized random forest. Calibration and comparisons between the highest and lowest predicted-effect groups do not consistently validate the predicted ordering. The resulting profiles describe the predictions of each learner, but do not establish which students would benefit most from the intervention.[54]

[52]Prior exposure to a female teacher could shape students' responses to female role models, but this comparison does not isolate that channel. Supplementary Table S35 extends the comparison to all survey domains and broad applications; no interaction survives the wild-bootstrap adjustment. Figures S9–S10 display the positive immediate belief difference, the near-zero follow-up difference, and the negative application point estimate. Teacher gender was not randomized, so differences across teaching contexts do not identify the effect of assigning a female teacher. Sparse application events and limited support in some assignment-by-teacher-gender cells constrain inference; failure to reject equal effects does not establish their equivalence.

[53]Supplementary Tables S38–S45 report the complete social-, time-, and risk-preference comparisons. The original multiple-testing families are preserved across the grouping variables.

[54]Supplementary Tables S46 and S47 report the calibration and group-effect tests. For identity, the LASSO's highest predicted-effect group has an estimated effect 0.316 standard deviations below the lowest group (S.E. = 0.151), a statistically significant difference; the forest contrast is 0.102 (S.E. = 0.211) and is not statistically significant. For applications, LASSO also reverses the ranking, whereas the forest's high-minus-low coefficient is 3.879 p.p. (S.E. = 3.104) and is not statistically significant. The LASSO uses school-level cross-fitting. The native forest uses observation-level folds for its exploratory ranking; its calibration and group-effect tests use the orthogonal score with school-clustered inference. Model-implied gender comparisons also depend on the learner: the forest predicts larger follow-up identity effects for women and smaller intentions and stereotype effects, while the corresponding LASSO gender differences are not statistically significant (Table S48). Tables S49–S53 associate baseline characteristics with predicted CATE scores. For applications, the forest profiles emphasize baseline application intentions, identity, and interest in economics, whereas the LASSO profiles emphasize age and classroom composition. Their p-values test associations with predicted scores, rather than treatment-effect differences identified by random assignment. Read together with the calibration results, they provide an exploratory description of the fitted models, without establishing reliable beneficiary profiles or causal mechanisms.

10 Robustness

10.1 Attrition corrections

Incomplete follow-up raises the concern that respondents differ across treatment arms. Online Appendix Table A25 shows a control response rate of 48.3 percent and a treated-school difference of 13.7 p.p. (S.E. = 7.0), which is statistically significant at 10 percent. The online-endline difference is smaller, at 3.1 p.p. (S.E. = 2.5) from a control mean of 22.8 percent, and is not statistically significant. We therefore examine whether reweighting respondents or bounding the effects changes the interpretation of the survey results.

Inverse-probability weighting preserves the positive belief effects at both horizons. The immediate complete-case estimate of 0.301 standard deviations becomes 0.297 under stabilized weights and 0.298 when weights are capped at the 99th percentile; all have p-value < 0.001 (Online Appendix Table A27). Immediate perceived fit also remains positive and statistically significant, at 0.088 under stabilized weights (S.E. = 0.022). At follow-up, weighting reduces the belief coefficient from 0.169 to 0.141 standard deviations (S.E. = 0.051), still statistically significant, with almost the same result after capping weights (Table A26). Thus, selection on the observed variables used to predict response does not account for the follow-up belief finding.[55]

Lee bounds give a more qualified conclusion under monotone sample selection (Online Appendix Table A28). The estimated interval for immediate beliefs is $[0.221, 0.351]$ standard deviations, and the bootstrapped confidence interval for the lower endpoint remains positive, at $[0.084, 0.357]$. This supports a positive immediate response under the bounding assumptions. The follow-up interval is $[-0.075, 0.431]$, which includes zero. The later belief result is therefore robust to reweighting on observed characteristics, but not established under the less restrictive Lee exercise.[56]

[55] Supplementary Table S54 shows that the school-assignment effect on follow-up response is 13.220 p.p. larger for women than men (wild-bootstrap q-value < 0.001), making selection relevant to comparisons by gender. In Supplementary Table S55, follow-up belief estimates with stabilized gender-specific weights are 0.193 standard deviations for women and 0.082 for men, compared with unweighted estimates of 0.194 and 0.127, respectively. The weighted women-minus-men difference remains statistically insignificant, at 0.111 (S.E. = 0.088; q-value = 0.346). Thus, the positive follow-up belief estimate for women is preserved, with no statistically significant gender difference. Weighting also preserves positive immediate belief effects of 0.250 for men and 0.341 for women, with no statistically significant difference between them. The immediate identity difference remains statistically significant at 10 percent after the conventional within-wave adjustment (q-value = 0.083 with stabilized weights and q-value = 0.051 with capped weights). Response models use assignment, blocks, baseline outcomes, and baseline characteristics. These estimates assess sensitivity to selection captured by those variables; they do not address selection on unobserved outcome determinants.

[56] The bounds trim the outcome distribution in the arm with the higher observation rate to equalize response

10.2 Alternative outcome definitions and specifications

The belief results do not depend on the selected precision covariates. Omitting both covariates and randomization-block fixed effects gives an immediate estimate of 0.284 standard deviations (S.E. = 0.058) and a follow-up coefficient of 0.177 (S.E. = 0.064); both are statistically significant (Online Appendix Table A30). Among specifications retaining block fixed effects and varying baseline-outcome and additional adjustment, immediate beliefs range from 0.269 to 0.302, with p-value < 0.001 throughout (Table A31). Follow-up beliefs range from 0.142 to 0.169 and remain statistically significant at our 10 percent threshold (Table A32). For women's applications, the expanded covariate set gives administrative and gateway estimates of 0.432 and 0.541 p.p., neither statistically significant (Table A33).[57]

The belief finding is more sensitive to aggregation. On a common complete sample of 2,230 students, the immediate inverse-covariance-weighted estimate is 0.286 standard deviations (S.E. = 0.051) and the equal-weight coefficient is 0.260 (S.E. = 0.063); both are statistically significant (Online Appendix Table A34). The first-principal-component estimate is 0.021 (S.E. = 0.075) and is not statistically significant. At follow-up, the common sample contains 1,287 students; ICW and equal-weight coefficients are 0.154 (S.E. = 0.060) and 0.148 (S.E. = 0.068), both statistically significant. The principal-component coefficient is 0.025 (S.E. = 0.067) and is not statistically significant (Table A35). The positive response thus survives equal weighting at both horizons, but does not extend to every index construction. These comparisons hold the respondent sample fixed within each wave, rather than mixing aggregation and sample-composition changes.

We find no statistically significant school-assignment effect on women's applications under the broader definitions. Women's effects are 0.404 p.p. under the administrative endpoint, 0.542 under

rates, using exact fractional trimming. The procedure assumes that assignment moves observation probabilities in the same direction for all students; observed response differences do not establish this restriction. Uncertainty uses 999 bootstrap replications that resample the original school assignment units within the original randomization blocks. The estimated bound interval is not a confidence interval. Complete-case ANCOVA remains the preregistered specification.

[57] Immediate identity is more sensitive to conditioning on its baseline measure. Its estimate is 0.012 standard deviations with block fixed effects alone (S.E. = 0.048), which is not statistically significant. The main ANCOVA coefficient of 0.097 (S.E. = 0.025) and expanded-control coefficient of 0.082 (S.E. = 0.021) are both statistically significant (Table A31). Comparisons on a common sample show that the change reflects adjustment for lower baseline perceived fit in treated schools. The block-adjusted baseline difference is −0.120 standard deviations. For broad applications, the comparison omitting block effects and all controls is 1.226 p.p. (S.E. = 0.633), statistically significant at 10 percent (Table A30), whereas the assignment-block coefficient is 0.542 (S.E. = 0.367) and is not statistically significant. These secondary results should not be described as invariant to specification.

the broad gateway, and 0.333 under the combined-source definition; the respective p-values are 0.148, 0.151, and 0.304 (Online Appendix Table A36). The all-student broad-gateway estimate is 0.321 (S.E. = 0.398) and is not statistically significant (Table A37). Thus, the absence of a statistically significant application effect also holds under these broader definitions.[58]

Finally, all estimates apply the common rule for the measurement problem described in Section 4.4: the 120 original records without a recoverable, comparable baseline are excluded, while valid observations from replacement classrooms remain (Online Appendix Table A2). Additionally excluding the replacement classrooms changes immediate beliefs from 0.301 to 0.278 standard deviations (S.E. = 0.062) and follow-up beliefs from 0.169 to 0.155 (S.E. = 0.054). Both coefficients remain statistically significant, so the positive response at either horizon does not depend on retaining those classrooms. Excluding replacement classrooms also leaves the estimated effect on women's administrative applications statistically insignificant, at 0.439 percentage points (S.E. = 0.301).[59]

10.3 Placebo tests and subject-specific benchmarks

Panel A of Online Appendix Table A39 examines whether the intervention produced a general increase in reported interest across unrelated subjects. The follow-up effects are −0.012 standard deviations for English, −0.030 for history, and −0.043 for Spanish; none is statistically distinguishable from zero. These placebo tests provide no pooled evidence of a generalized subject-interest response, but cannot rule out selective reporting effects or resolve demand at the immediate survey.[60]

[58] The secondary institution-specific signal remains relevant: applications to UANL among all students increase under school assignment and direct treatment, as discussed in Section 7.3. Women's broad-gateway T–C estimate is 0.775 p.p. (S.E. = 0.399), statistically significant under conventional inference at 10 percent, but does not pass the wild-bootstrap or randomization tests at 10 percent (p-value = 0.102 and p-value = 0.143). The gateway definition changes program coverage, while the combined-source measure also uses August-survey reports. Together, these comparisons assess whether the findings depend on program scope or the sources used to observe educational choices.

[59] Immediate identity remains positive after the exclusion, at 0.075 standard deviations (S.E. = 0.027), and remains statistically significant, while effects on immediate intentions and stereotypes remain statistically insignificant. These sensitivity tests use conventional school-clustered inference without additional multiplicity adjustment.

[60] Supplementary Table S56 tests whether these pooled estimates conceal differences between women and men. The treated-school assignment effect on interest in history is 0.135 standard deviations lower for women (S.E. = 0.056); the direct-treatment-versus-control coefficient is similar, at −0.159 (S.E. = 0.062). Both differences are statistically significant at 10 percent after adjustment (q-value = 0.070 and q-value = 0.051, respectively). School-assignment gender differences for English and Spanish are not statistically significant, nor are those for the mathematics and economics outcomes. Thus, the pooled results do not imply that all gender differences are zero, although the history contrasts alone do not identify a reporting mechanism. These exploratory comparisons use conventional clustered inference. The reported q-values apply the Benjamini–Hochberg adjustment separately within each outcome family

Panels B and C report complementary outcomes for mathematics and economics: subject interest and self-assessed ability. We find no statistically significant effects on these outcomes. These measures may respond to the intervention and are interpreted as related and targeted outcomes, respectively, rather than as placebo tests.

11 Cost-Effectiveness

Including visits to direct-treatment, spillover, and control classrooms, we estimate that delivering the classroom sessions cost USD 13,450–15,550, or USD 3.57–4.13 per student who completed the baseline survey. This is lower than the USD 5 per student that Riley (2024) reports for the film screening and transportation in Uganda, but higher than the USD 0.86 per student that Asanov et al. (2026) report for delivering role-model videos at scale in Ecuador.[61] Female economists delivered the role-model sessions, while student presenters delivered the active placebo in spillover and control classrooms. The range reflects the number of groups served on each trip. With six study groups per campus, visiting three groups at a time requires 60 round trips across the 30 campuses. Visiting two at a time requires 90.

Project management, including preparation of materials and training, cost USD 5,000. A second coordinator managed the fieldwork for USD 3,000. Production of the three videos cost USD 750, and the server cost USD 500 over the preparation and implementation period. We budget USD 4,200–6,300 for transportation and meals across all initial visits.[62] Classroom teachers administered the follow-up survey, and the laboratory team monitored it remotely. Adding USD 4,000 for the ten tablets offered as survey incentives and USD 4,687 for printing supplied by the Tec brings the study budget to USD 22,137–24,237, or USD 5.88–6.44 per baseline respondent.[63]

and assignment contrast, rather than jointly across all 28 comparisons.

[61] The 3,763 baseline respondents comprise 919 students in directly treated classrooms, 939 in spillover classrooms, and 1,905 in control classrooms (Table 2). We use baseline completion to measure reach because individual attendance was not recorded for all sessions. Riley budgets USD 3 for admission and USD 2 for transportation per student attending the screening. We compare costs at the prices reported in each study. These comparisons are illustrative, as costing conventions differ across studies.

[62] We use a planning exchange rate of MXN 20 per U.S. dollar. A round trip costs USD 50, and food costs USD 5 per person per trip. We budget for two assistants, a presenter, and one coordinator.

[63] At USD 0.05 per single-sided sheet, a reserve of paper questionnaires for 30 percent of the 5,859 rostered students costs USD 1,757.70. The baseline, immediate, follow-up, and August forms have 5, 6, 6, and 3 pages, respectively. The reserve includes the August instrument, which students completed online. The field team printed long parental forms, available in six- and seven-page versions. The number distributed alongside the short parental notice is unknown. We budget USD 2,929.50 for one seven-page parental form, a two-page student consent form, and a one-page notice per rostered student.

The role models and student assistants contributed their time to the visits. Valuing professional time at USD 50 per hour and student time at USD 5 per hour raises the estimated cost to USD 28,067–31,217 when we include the visiting economists and two assistants.[64]

For a school that already offers student-led career talks, we can also assess the additional cost of offering the role-model session. Holding the itinerary and coordination fixed, we estimate an additional USD 2.85–3.32 per student in the 43 directly assigned classrooms. This amount covers the three videos and the difference in the value of the economist's and the student presenter's time. Relative to control classrooms, the role-model session increased the immediate beliefs index by 0.579 standard deviations (Online Appendix Table A18). At that effect size, the additional cost is USD 0.49–0.57 per student for a 0.1-standard-deviation improvement in beliefs.[65] We find no statistically significant increase in women's applications to economics, so the experiment does not yield a reliable cost per additional female applicant.

12 Conclusion

A narrow view of what economists do may discourage students from considering economics when they choose a university major. To examine this possibility, we conducted a field experiment in high schools in Nuevo León, Mexico, comparing a visit by a female economist with a student-led career talk of the same duration. Relative to the control group, assignment to a treatment school increased the beliefs-about-economics index by 0.30 standard deviations immediately after the visit and by 0.17 several weeks later. However, we find no statistically significant increase in women's applications to economics.

Immediately after the visit, students in treatment schools reported a better understanding of economists' work and were less likely to view economics as confined to money and finance. At follow-up, item-level evidence was clearest for rejection of the money-and-finance stereotype. The

[64] We allow one hour per assigned classroom for each assistant, one hour for each of the 43 role-model sessions, and an hour of travel per round trip. We allocate an economist's travel time to each of the 30–45 trips to treated campuses. This puts the value of the economists' donated time, including travel, at USD 85–102 per talk. For comparison, Agurto et al. (forthcoming) report in the working-paper version of their Peruvian study that presenters received USD 230 on average and gave an average of five talks. This implies about USD 46 in compensation per talk, with payments based on travel distance. Additional hours for unpaid training, any separate student presenter in placebo classrooms, teacher administration of the follow-up, laboratory monitoring, and survey processing remain unpriced where they extend beyond paid coordination time.

[65] We divide the additional resources required for the 43 classrooms by their 1,417 rostered students. This calculation assumes that the effect estimated among survey respondents applies to the full classroom roster.

immediate gains in beliefs were positive for both women and men and concentrated in directly treated classrooms. Our framework of motivated beliefs distinguishes how students understand a profession from whether they see a place for themselves in it. Although students reported a greater sense of fit with economics immediately after the visit, we find no statistically detectable follow-up effects on the same item, the broader self-concept index, intentions and aspirations, or gender stereotypes. This pattern is consistent with students revising their beliefs about the field while their own career plans changed little.

We do find a small but statistically significant increase in applications to economics at UANL when we pool women and men. In Peru, where students faced a similar choice, Agurto et al. (forthcoming) find that visits by female engineers increased engineering applications and enrollment among girls with high mathematics grades in the partner university's main recruitment area. In the United States, Porter and Serra (2020) find that visits by alumnae led more women already taking introductory economics to major in the subject. Students' familiarity with a field when they meet a role model may therefore matter for how they respond.

For schools that already offer student-led career talks, we estimate that inviting an economist costs an additional USD 2.85–3.32 per student on the class roster. At that cost, schools can give students a broader understanding of economics before they decide where to apply. Future research could test whether pairing the visit with help identifying economics programs and completing the application when admissions open encourages more young women to apply.

Use of artificial intelligence

We used AI tools to improve the clarity of the manuscript, to prepare the replication materials, and to correct errors in English grammar and usage.

References

Agurto, Marcos, Muchin Bazan, Siddharth Hari, and Sudipta Sarangi. forthcoming. "To Inspire and to Inform: The Role of Role Models." *The World Bank Economic Review.*

Agurto, Marcos, Sandra Boisvert, Siddharth Hari, Valeria Quevedo, Sudipta Sarangi, and Susana Vegas. 2025. "Signals of Aptitude and Female STEM Career Choices." *Economics of Education Review*, 109: 102727.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–1495.

Antman, Francisca M., Evelyn Skoy, and Nicholas E. Flores. 2026. "Can Better Information Reduce College Gender Gaps? The Impact of Relative Grade Signals on Academic Outcomes for Students in Introductory Economics." *Southern Economic Journal*, 93(1): 200–239.

Arceo-Gómez, Eva O. 2025. "Status of Women in Economics: Mexico." Centro de Investigación y Docencia Económicas (CIDE) Documento de Trabajo DTE 649, Mexico City.

Asanov, Igor, Thomas Åstebro, Guido Buenstorf, Bruno Crépon, Francisco Patricio Flores-Taipe, David McKenzie, Mona Mensmann, and Mathis Schulte. 2026. "Remote Delivery of STEM and Entrepreneurship Role Models at Scale Changes College Major Choice in Ecuador." *Nature Human Behaviour*, 10: 1223–1233.

Avery, Mallory, Jane Caldwell, Christian D. Schunn, and Katherine Wolfe. 2024. "Improving Introductory Economics Course Content and Delivery Improves Outcomes for Women." *The Journal of Economic Education*, 55(3): 216–231.

Avilova, Tatyana, and Claudia Goldin. 2018. "What Can UWE Do for Economics?" *AEA Papers and Proceedings*, 108: 186–190.

Avilova, Tatyana, and Claudia Goldin. 2024. "Seeking the 'Missing Women' of Economics with the Undergraduate Women in Economics Challenge." *Journal of Economic Perspectives*, 38(3): 137–162.

Bansak, Cynthia, and Martha Starr. 2010. "Gender Differences in Predispositions towards Economics." *Eastern Economic Journal*, 36(1): 33–57.

Barrera, Sergio, Susan Sajadi, Marionette Holmes, and Sarah Jacobson. 2024. "Valuing Identity in the Classroom: What Economics Can Learn from Science, Technology, Engineering, and Mathematics Education." *Journal of Economic Perspectives*, 38(3): 163–190.

Bayer, Amanda, and Cecilia Elena Rouse. 2016. "Diversity in the Economics Profession: A

New Attack on an Old Problem." *Journal of Economic Perspectives*, 30(4): 221–242.

Bayer, Amanda, Syon P. Bhanot, and Fernando Lozano. 2019. "Does Simple Information Provision Lead to More Diverse Classrooms? Evidence from a Field Experiment on Undergraduate Economics." *AEA Papers and Proceedings*, 109: 110–114.

Bedard, Kelly, Jacqueline Dodd, and Shelly Lundberg. 2021. "Can Positive Feedback Encourage Female and Minority Undergraduates into Economics?" *AEA Papers and Proceedings*, 111: 128–132.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics*, 117(3): 871–915.

Bénabou, Roland, and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives*, 30(3): 141–164.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1): 289–300.

Bertocchi, Graziella, Luca Bonacini, and Marina Murat. 2023. "Adams and Eves: High School Math and the Gender Gap in Economics Majors." *Economic Inquiry*, 61(4): 798–817.

Borges, Bruna, and Fernanda Estevan. 2024. "Does Exposure to More Women in Male-Dominated Fields Render Female Students More Career-Oriented?" *Management Science*, 71(6): 5045–5082.

Breda, Thomas, Julien Grenet, Marion Monnet, and Clémentine Van Effenterre. 2023. "How Effective Are Female Role Models in Steering Girls Towards STEM? Evidence from French High Schools." *The Economic Journal*, 133(653): 1773–1809.

Buckles, Kasey. 2019. "Fixing the Leaky Pipeline: Strategies for Making Economics Work for Women at Every Stage." *Journal of Economic Perspectives*, 33(1): 43–60.

Buzard, Kristy, Katelyn Cranney, Laura K. Gee, and Olga Stoddard. 2025. "What Deters Women from Economics." *AEA Papers and Proceedings*, 115: 659–663.

Chambers, Andrea, Stacy Dickert-Conlin, Carey Elder, Steven J. Haider, and Scott Imberman. 2021. "Info.Econ: Increasing Diversity among Economics Majors." *AEA Papers and Proceedings*, 111: 133–137.

Christensen, Johan. 2018. "Economic Knowledge and the Scientization of Policy Advice." *Policy Sciences*, 51(3): 291–311.

Christensen, Johan, and Ronen Mandelkern. 2022. "The Technocratic Tendencies of Economists in Government Bureaucracy." *Governance*, 35(1): 233–257.

Del Carpio, Lucía, and Maria Guadalupe. 2022. "More Women in Tech? Evidence from a Field Experiment Addressing Social Identity." *Management Science*, 68(5): 3196–3218.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." *American Economic Review*, 108(11): 3266–3302.

Diario Oficial de la Federación. 2012. "Decreto por el que se Declara Reformado el Párrafo Primero; el Inciso C) de la Fracción II y la Fracción V del Artículo 3o., y la Fracción I del Artículo 31 de la Constitución Política de los Estados Unidos Mexicanos." *Diario Oficial de la Federación, February 9*, https://www.diputados.gob.mx/LeyesBiblio/ref/dof/CPEUM_ref_201_09feb12.pdf.

Di, Aurelia. 2025. "The Power of Words: How Role Models Influence Belief and Behavior." SSRN working paper.

Dynan, Karen E., and Cecilia Elena Rouse. 1997. "The Underrepresentation of Women in Economics: A Study of Undergraduate Economics Students." *The Journal of Economic Education*, 28(4): 350–368.

Ersoy, Fulya, and Jamin D. Speer. 2025. "Opening the Black Box of College Major Choice: Evidence from an Information Intervention." *Journal of Economic Behavior & Organization*, 231: 106800.

Gentry, Melissa, Jonathan Meer, and Danila Serra. 2023. "Can High School Counselors Help the Economics Pipeline?" *AEA Papers and Proceedings*, 113: 462–466.

Ginther, Donna K., and Shulamit Kahn. 2004. "Women in Economics: Moving Up or Falling Off the Academic Career Ladder?" *Journal of Economic Perspectives*, 18(3): 193–214.

Ginther, Donna K., Shulamit Kahn, and Jessica McCloskey. 2016. "Gender and Academics." In *The New Palgrave Dictionary of Economics*, ed. Steven N. Durlauf and Lawrence E. Blume, 1–18. Palgrave Macmillan.

Golman, Russell, David Hagmann, and George Loewenstein. 2017. "Information Avoid-

ance." *Journal of Economic Literature*, 55(1): 96–135.

Halim, Daniel, Elizabeth T. Powers, and Rebecca Thornton. 2022. "Gender Differences in Economics Course-Taking and Majoring: Findings from an RCT." *AEA Papers and Proceedings*, 112: 597–602.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological Methods*, 15(4): 309–334.

Kahn, Shulamit. 1993. "Gender Differences in Academic Career Paths of Economists." *American Economic Review*, 83(2): 52–56.

Kahn, Shulamit B. 1995. "Women in the Economics Profession." *Journal of Economic Perspectives*, 9(4): 193–206.

Li, Hsueh-Hsiang. 2018. "Do Mentoring, Information, and Nudge Reduce the Gender Gap in Economics Majors?" *Economics of Education Review*, 64: 165–183.

Lundberg, Shelly, and Jenna Stearns. 2019. "Women in Economics: Stalled Progress." *Journal of Economic Perspectives*, 33(1): 3–22.

McKenzie, David. 2012. "Beyond Baseline and Follow-Up: The Case for More T in Experiments." *Journal of Development Economics*, 99(2): 210–221.

Owen, Ann L. 2010. "Grades, Gender, and Encouragement: A Regression Discontinuity Analysis." *The Journal of Economic Education*, 41(3): 217–234.

Owen, Ann L., and Paul Hagstrom. 2021. "Broadening Perceptions of Economics in a New Introductory Economics Sequence." *The Journal of Economic Education*, 52(3): 175–191.

Patnaik, Arpita, Gwyn C. Pauley, Joanna Venator, and Matthew J. Wiswall. 2024. "The Impacts of Same and Opposite Gender Alumni Speakers on Interest in Economics." *Economics of Education Review*, 102: 102579.

Porter, Catherine, and Danila Serra. 2020. "Gender Differences in the Choice of Major: The Importance of Female Role Models." *American Economic Journal: Applied Economics*, 12(3): 226–254.

Pugatch, Todd, and Elizabeth Schroeder. 2021. "Promoting Female Interest in Economics: Limits to Nudges." *AEA Papers and Proceedings*, 111: 123–127.

Riley, Emma. 2024. "Role Models in Movies: The Impact of Queen of Katwe on Students' Educational Attainment." *The Review of Economics and Statistics*, 106(2): 334–351.

Secretaría de Educación Pública. 2017. "Oferta Educativa: Educación Media Superior." *Secretaría de Educación Pública*, https://www.gob.mx/sep/es/acciones-y-programas/oferta-educativa-educacion-media-superior (accessed September 27, 2026).

Secretaría de Educación Pública. 2024. "Principales Cifras del Sistema Educativo Nacional 2023–2024." Secretaría de Educación Pública, Mexico City.

Serra, Danila. 2025. "Role Models in Developing Countries." In *Handbook of Experimental Development Economics*, ed. Utteeyo Dasgupta and Pushkar Maitra, Chapter 7, 123–135. Cheltenham, UK: Edward Elgar Publishing.

Smith, Sarah, Arun Advani, and Laura Harvey. 2025. "Going Back to the High Schools: Challenging Stereotypes about Economics." Centre for Economic Policy Research CEPR Discussion Paper 20483.

Steele, Claude M. 1997. "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance." *American Psychologist*, 52(6): 613–629.

WELAC. 2023. "The Status of Women in the Economics Profession in Latin America and the Caribbean." Women Economists in Latin America and the Caribbean (WELAC).

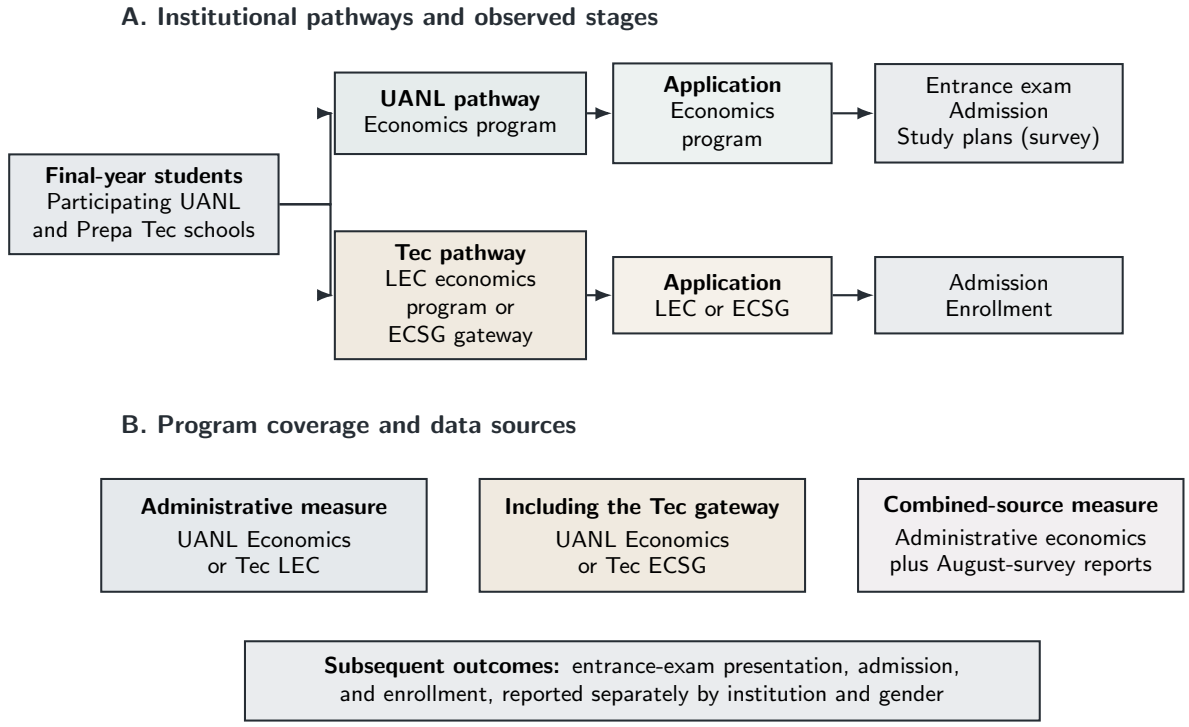


Figure 1: Educational Pathways and Economics Application Outcomes

Notes: Panel A maps the educational pathways observed for final-semester students in participating UANL and Prepa Tec schools. The UANL registry identifies applications to Economics, entrance-exam participation, and admission; it does not establish enrollment. The Tec records distinguish the LEC economics program from the broader ECSG gateway and identify admission and enrollment. August-survey responses provide additional information on students' reported study choices at these and other institutions. Panel B distinguishes program and source coverage. The administrative application measure identifies UANL Economics or Tec LEC. The gateway measure combines UANL Economics with Tec ECSG, which covers several disciplines. The combined-source measure uses the administrative economics indicator and August-survey economics reports, including cases absent from the registries. Administrative records and survey responses provide complementary coverage of educational choices at the study institutions and elsewhere. Applications, admission, and enrollment remain distinct events.

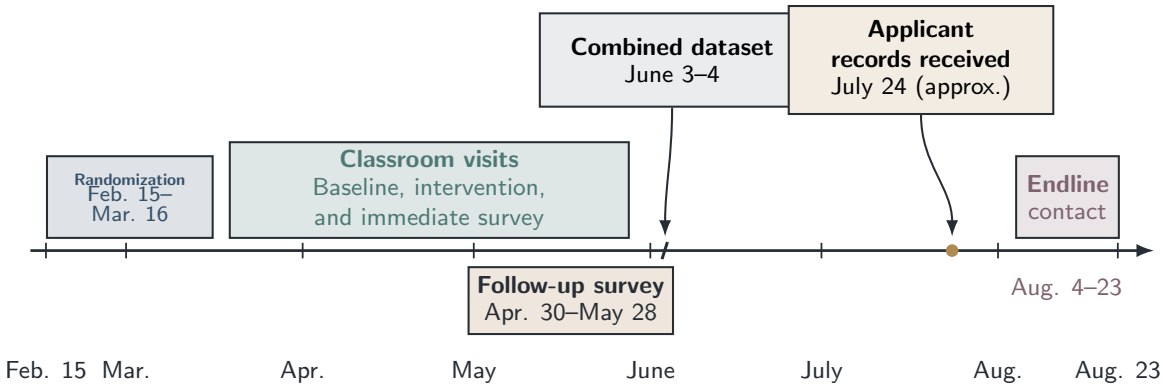


Figure 2: Experimental Design and Data-Collection Timeline

Notes: The figure places assignment, implementation, and data collection on a common 2026 calendar. Schools were assigned within the five original randomization blocks and classrooms within treated schools were subsequently assigned to direct treatment or spillover before student consent. During each classroom visit, the field team obtained consent, administered the baseline survey, delivered the assigned role-model or active-placebo activity, and administered the immediate post-intervention survey. The follow-up survey was fielded from April 30 through May 28. The combined survey database was assembled on June 3–4; administrative applicant records were received around July 24; and August online-survey contact occurred from August 4 through August 23. Horizontal segments show observed implementation windows, whereas isolated markers denote short operational events or approximate receipt dates. All causal estimates in the paper use the 2026 experimental cohort.

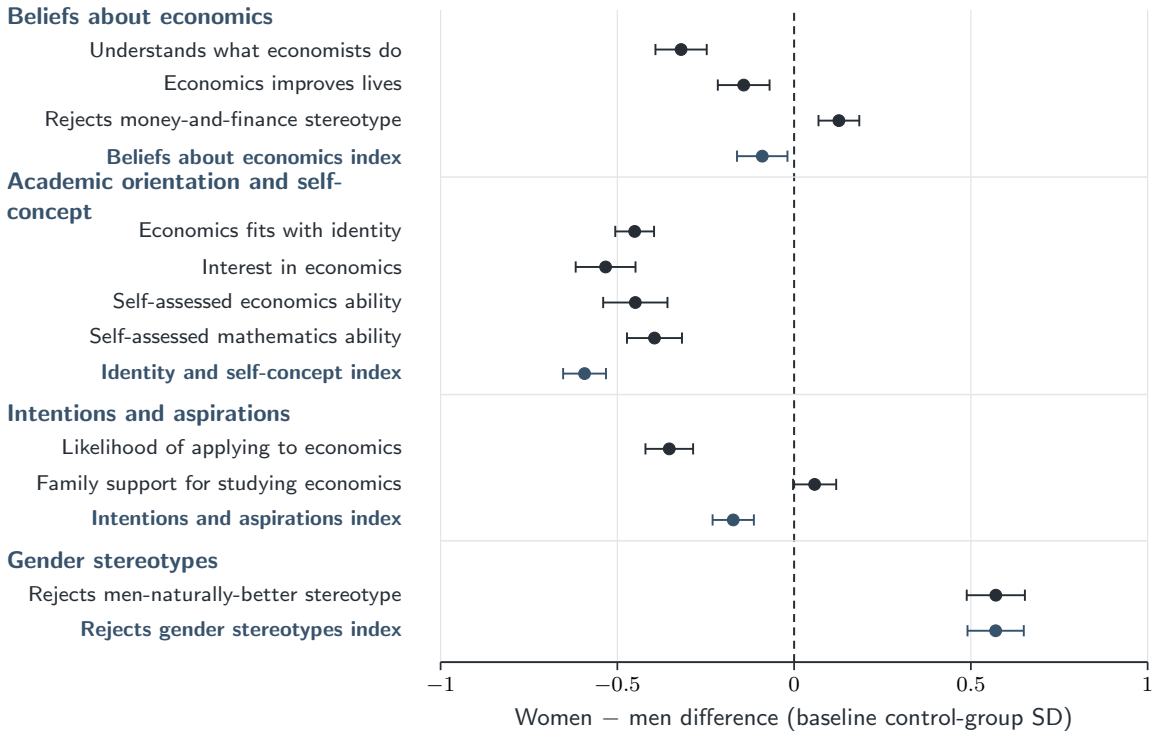


Figure 3: Baseline Gender Gaps

Notes: The sample consists of final-semester baseline respondents in control schools, so the plotted gaps are not affected by exposure to the intervention. Dots report women-minus-men coefficients from regressions containing the original randomization-block fixed effects; horizontal bars are 95 percent confidence intervals based on school-clustered standard errors. Items and indices are grouped into beliefs about economics, academic orientation and self-concept, intentions and aspirations, and gender stereotypes. Outcomes are standardized using baseline control-group moments and oriented so that positive values indicate more economics-oriented responses or less stereotyped views. A point to the left of zero therefore indicates a lower mean among women under this common orientation. The dashed vertical line denotes equality between women and men. These are descriptive baseline associations and are not interpreted causally.

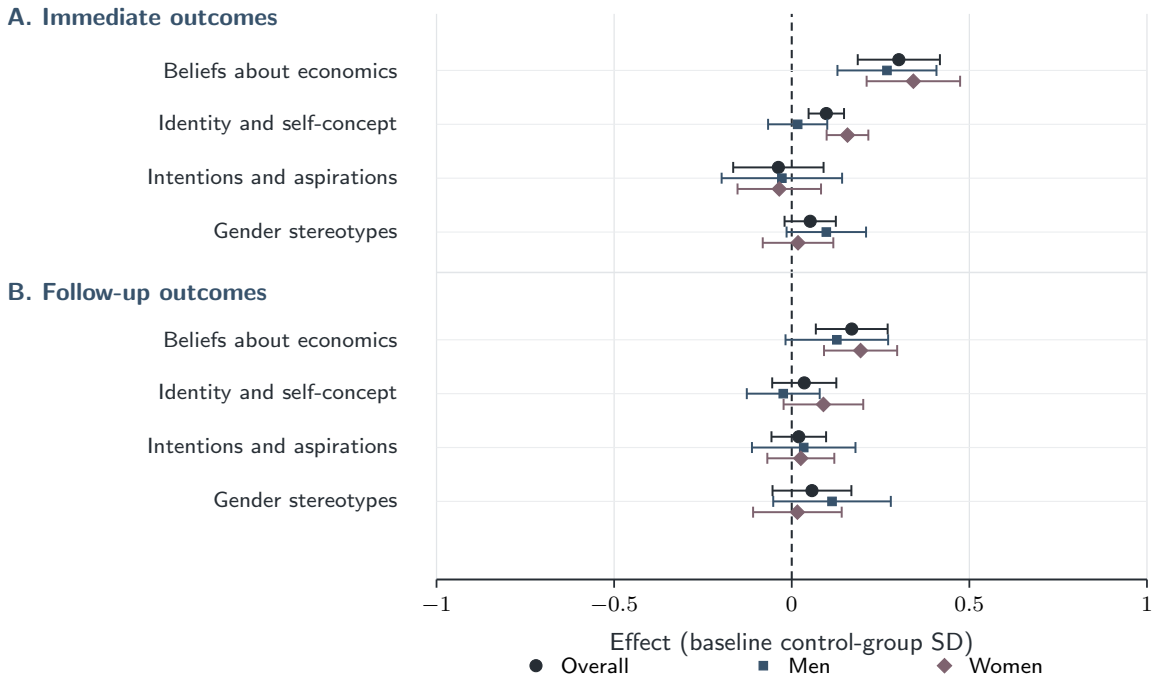


Figure 4: PAP Index Effects Overall and by Gender

Notes: Panel A reports immediate post-intervention outcomes and Panel B follow-up outcomes. Circles show the overall treated-school assignment effect, squares the effect for men, and diamonds the effect for women; horizontal bars are 95 percent confidence intervals and the dashed vertical line denotes zero. Treated-school assignment combines directly treated and spillover classrooms in treated schools and compares them with classrooms in control schools. Overall estimates use the preregistered ANCOVA. Gender-specific estimates are linear combinations from a pooled treatment-by-gender interaction model, rather than separate regressions. Every specification includes the original randomization-block fixed effects and the index's baseline value and uses school-clustered standard errors. Outcomes are standardized using baseline control-group moments and oriented so that positive effects indicate more economics-oriented responses or less stereotyped views. Both panels use the common horizontal scale from −1 to 1.

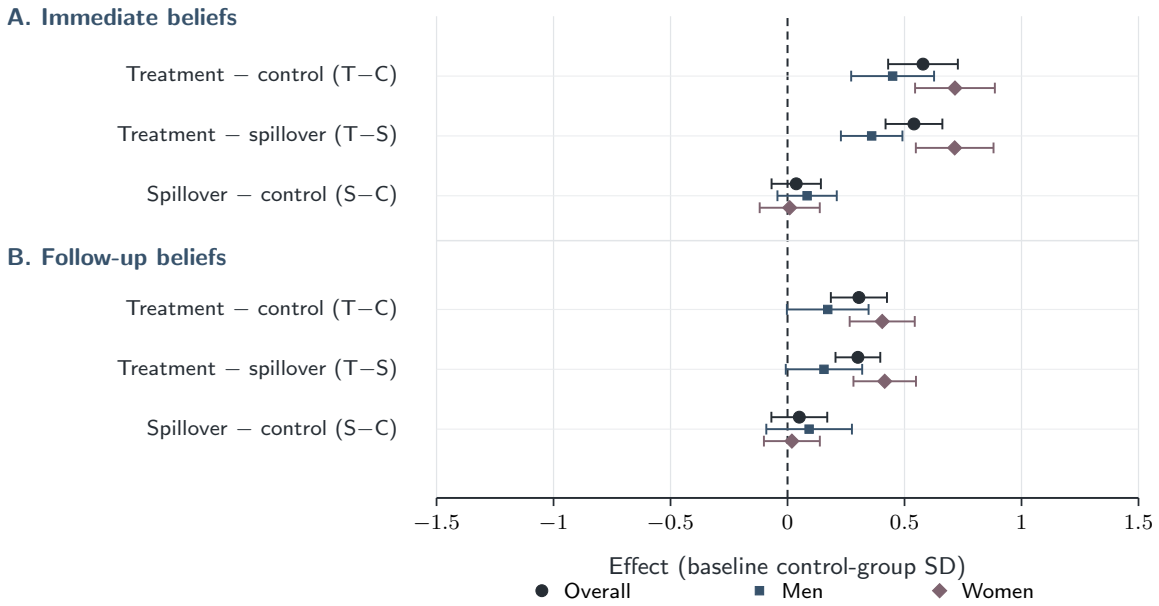


Figure 5: Direct and Spillover Effects on Beliefs Overall and by Gender

Notes: Panel A reports immediate beliefs and Panel B follow-up beliefs. Circles show overall effects, squares effects for men, and diamonds effects for women; horizontal bars are 95 percent confidence intervals and the dashed vertical line denotes zero. Direct treatment versus control compares direct-treatment classrooms with classrooms in control schools, and spillover versus control compares spillover classrooms in treated schools with classrooms in control schools. These cross-school contrasts include the original randomization-block fixed effects and use school-clustered standard errors. Direct treatment versus spillover is identified within treated schools and includes school fixed effects with classroom-clustered standard errors. Gender-specific estimates are obtained from pooled interaction models. Beliefs are standardized using baseline control-group moments and oriented so that positive effects indicate more economics-oriented responses. Both panels use the common horizontal scale from −1.5 to 1.5, which displays every point estimate and confidence interval without truncation.

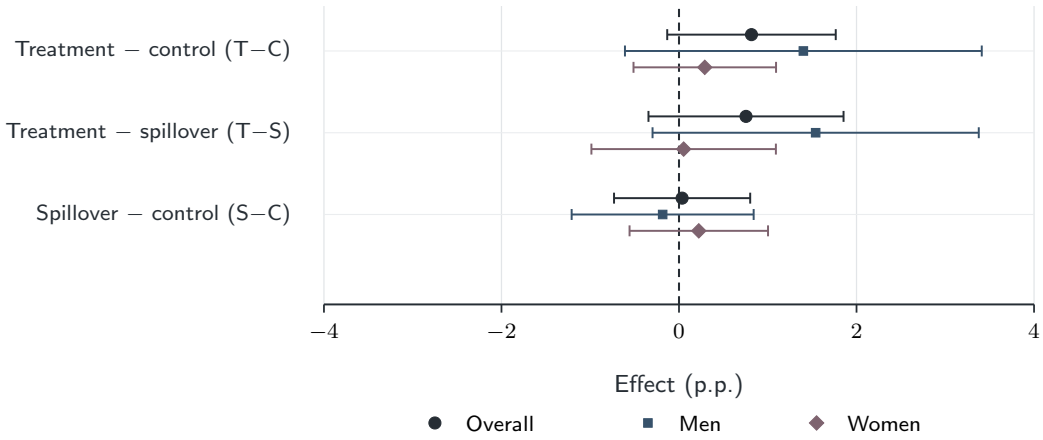


Figure 6: Direct and Spillover Effects on Administrative Economics Applications Overall and by Gender

Notes: The outcome is the administrative economics-application indicator, equal to one for an application to UANL Economics or Tec LEC. Circles show overall effects, squares effects for men, and diamonds effects for women; horizontal bars are 95 percent confidence intervals and the dashed vertical line denotes zero. Direct treatment versus control compares direct-treatment classrooms with classrooms in control schools, and spillover versus control compares spillover classrooms in treated schools with classrooms in control schools. These cross-school contrasts include the original randomization-block fixed effects and use school-clustered standard errors. Direct treatment versus spillover is identified within treated schools and includes school fixed effects with classroom-clustered standard errors. Gender-specific estimates come from pooled interaction models. The binary outcome is multiplied by 100, so all effects and confidence intervals are in percentage points (p.p.). The common horizontal scale runs from −4 to 4 p.p.

Table 1: Experimental Sample and Preassignment School Balance

Sample or characteristic	Control schools (1)	Treated schools (2)	Total sample size (3)	Block-adjusted difference (4)	Adjusted standardized difference (5)
Panel A. Randomization frame and realized baseline sample					
Classrooms	83	85	168		
Rostered students	3,045	2,814	5,859		
Baseline respondents	1,905	1,858	3,763		
Baseline participation rate	62.6%	66.0%			
Panel B. Preassignment school-unit characteristics, UANL and Tec					
Eligible Proyecto de Vida classrooms	24.286	21.133	29	-2.020	-0.149
Students in eligible classrooms	953.786	763.667	29	-128.877	-0.208
Mean eligible-class size	36.169	34.209	29	-1.102	-0.118
Female share of eligible students	0.529	0.523	29	-0.010	-0.146
Share of eligible classes taught by women	0.801	0.770	29	-0.037	-0.178
Share of eligible classes in morning shift	0.422	0.602	29	0.154	0.380
Panel C. Additional school characteristics assembled in 2025, UANL					
Total school enrollment in 2025	2810.634	2540.083	24	-159.109	-0.082
Proyecto de Vida groups in 2025	22.909	20.000	23	-1.982	-0.153
Proyecto de Vida instructors in 2025	4.545	5.250	23	0.995	0.317
Female share of Proyecto de Vida instructors	0.779	0.797	23	0.004	0.016

Notes: Panel A summarizes the experimental roster and realized baseline survey completion across the UANL and Tec school networks. Treated schools include both directly treated and spillover classrooms. Panels B and C use one observation per original school-randomization unit and do not condition on student consent, attendance, or survey response. Panel B uses the preassignment Proyecto de Vida classroom rosters for all 29 randomized units. Panel C uses additional school information assembled in 2025 for the 24 UANL units; group and instructor measures are unavailable for one unit. Control and treated means are unadjusted. The block-adjusted difference is the coefficient on treated-school assignment from a regression with the original randomization-block fixed effects. The adjusted standardized difference divides this coefficient by the pooled within-arm standard deviation. Consistent with common balance diagnostics, the table emphasizes magnitude rather than hypothesis tests; absolute standardized differences near or above 0.25 warrant attention. Total sample size counts classrooms or students in Panel A, as indicated by the row, and school assignment units in Panels B and C.

Table 2: Classroom Assignment and Survey Completion

Assignment arm	Classrooms	Rostered students	Baseline responses	Immediate responses	Follow-up responses	Online endline responses
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. All schools						
Control	83	3,045	1,905 (62.6%)	1,849 (60.7%)	1,247 (41.0%)	446 (14.6%)
Spillover	42	1,397	939 (67.2%)	904 (64.7%)	712 (51.0%)	298 (21.3%)
Direct treatment	43	1,417	919 (64.9%)	901 (63.6%)	808 (57.0%)	257 (18.1%)
Panel B. Tec school network						
Control	12	322	49 (15.2%)	43 (13.4%)	18 (5.6%)	7 (2.2%)
Spillover	9	236	145 (61.4%)	138 (58.5%)	62 (26.3%)	67 (28.4%)
Direct treatment	9	259	166 (64.1%)	164 (63.3%)	67 (25.9%)	84 (32.4%)
Panel C. UANL school network						
Control	71	2,723	1,856 (68.2%)	1,806 (66.3%)	1,229 (45.1%)	439 (16.1%)
Spillover	33	1,161	794 (68.4%)	766 (66.0%)	650 (56.0%)	231 (19.9%)
Direct treatment	34	1,158	753 (65.0%)	737 (63.6%)	741 (64.0%)	173 (14.9%)

Notes: Panel A pools the Tec and UANL school networks; Panels B and C report the corresponding network-specific counts. Classroom and enrollment figures refer to all assigned classrooms, including classrooms in which a visit was not completed, whereas survey figures are realized responses. Completion rates use the assignment roster as the denominator and therefore combine absence, refusal, and non-completion; they should not be interpreted as attendance or consent rates. Spillover classrooms did not receive a role-model visit but remained eligible for the study surveys. The rescheduled treatment visit to classroom 82 is included. The confirmed roster transcription error for classroom 108 is corrected from 409 to 40 students. Survey columns report response counts, with completion rates in parentheses.

Table 3: Effects on Student Outcomes, by Gender and Survey Wave

Outcome	Women: control mean (1)	Women: effect (2)	Men: control mean (3)	Men: effect (4)	Women - men (5)	p-value [q-value] (difference) (6)
Panel A. Follow-up survey						
Beliefs about economics	-0.063	0.194***	-0.083	0.127*	0.067	0.362 [0.483]
		(0.053)		(0.074)	(0.072)	
Identity and self-concept	-0.420	0.089	0.248	-0.024	0.113*	0.078 [0.314]
		(0.057)		(0.052)	(0.062)	
Intentions and aspirations	-0.410	0.026	-0.247	0.034	-0.008	0.931 [0.931]
		(0.048)		(0.074)	(0.094)	
Gender stereotypes	0.079	0.016	-0.442	0.113	-0.097	0.285 [0.483]
		(0.064)		(0.084)	(0.089)	
Panel B. Immediate survey						
Beliefs about economics	0.094	0.342***	0.113	0.268***	0.074	0.289 [0.446]
		(0.067)		(0.071)	(0.069)	
Identity and self-concept	-0.246	0.157***	0.274	0.017	0.140**	0.011 [0.044]
		(0.030)		(0.042)	(0.051)	
Intentions and aspirations	-0.196	-0.035	-0.041	-0.028	-0.007	0.913 [0.913]
		(0.060)		(0.087)	(0.066)	
Gender stereotypes	0.220	0.018	-0.320	0.097*	-0.080	0.335 [0.446]
		(0.051)		(0.057)	(0.081)	

Notes: Panel A reports follow-up outcomes and Panel B immediate post-intervention outcomes. The treated-school effect compares all classrooms in treated schools, whether directly treated or assigned to spillover, with classrooms in control schools. Men and women estimates are obtained from a single pooled treatment-by-gender interaction model for each outcome; Women–men is the corresponding interaction coefficient. All specifications include the original randomization-block fixed effects and the outcome's baseline value and use school-clustered standard errors. Survey-index effects and control means are in standard deviations relative to the baseline control group; positive values indicate more economics-oriented responses or less stereotyped views. Within each panel, the reported q-value controls the false-discovery rate across the four displayed women–men interaction tests. Gender-specific p-values and q-values are retained in the replication files. The final column reports the women-minus-men interaction p-value, with its q-value in square brackets. Standard errors (S.E.) are reported in parentheses. *, **, and *** denote two-sided conventional p-values below 0.10, 0.05, and 0.01, respectively. Significance is assessed at 10 percent using unrounded p-values; adjusted tests use q-values below 0.10.

Table 4: Effects on Applications to Economics

	Combined measure (registry or self-report) (1)	Administrative measure (UANL or Tec registry) (2)	UANL only (3)	Tec only (4)	Other (self-rep. only) (5)
Panel A. Women (PAP primary sample)					
PAP specification	0.333	0.404	0.404	0.031	0.031
	(0.318)	(0.271)	(0.271)	(0.036)	(0.036)
Wild-bootstrap p-value	0.361	0.182	0.187	0.377	0.386
Randomization-inference p-value	0.371	0.211	0.211	0.895	0.895
Covariate-adjusted specification	0.340	0.432	0.432	0.024	0.024
	(0.329)	(0.275)	(0.275)	(0.033)	(0.033)
Control mean (percent)	0.294	0.196	0.196	0.000	0.000
Sample size	2,009	2,009	2,009	2,009	2,009
Number of schools	28	28	28	28	28
Panel B. All final-semester students					
PAP specification	0.490	0.412	0.659**	-0.193	0.096
	(0.387)	(0.325)	(0.254)	(0.164)	(0.058)
Wild-bootstrap p-value	0.280	0.280	0.020	0.414	0.198
Randomization-inference p-value	0.218	0.232	0.039	0.162	0.201
Covariate-adjusted specification	0.614	0.516	0.741***	-0.166	0.112*
	(0.371)	(0.309)	(0.251)	(0.150)	(0.066)
Control mean (percent)	0.630	0.472	0.315	0.157	0.000
Sample size	3,763	3,763	3,763	3,763	3,763
Number of schools	28	28	28	28	28

Notes: Panel A reports women in the final semester, the pre-analysis-plan primary population for the administrative outcome; Panel B reports all final-semester students. The treated-school coefficient compares students in treated schools with students in control schools within each panel's sample and includes the five original randomization-block fixed effects; standard errors are clustered at the school assignment unit. Application outcomes are binary indicators multiplied by 100. Effects and standard errors are reported in percentage points (p.p.); control means are reported as percentages. A one-percentage-point effect corresponds to a change of 0.01 in probability units. The combined measure equals one when an economics application is found in the UANL registry, the Tec registry, or the endline self-report. The administrative measure equals one for an application to UANL Economics or Tec LEC; UANL and Tec report the institution-specific components. Other records self-reported economics applications outside both registries. The combined measure and the Other category record detected events in the full eligible sample; no survey-response indicator enters the regressions. A zero records no detected event, not verified nonparticipation. Differences in August-survey coverage can therefore affect event detection in these measures. Adjusted models add baseline age, GPA, application propensity, interest in economics, and self-assessed mathematics ability, with mean imputation and missingness indicators. Wild-bootstrap p-values use a null-imposed school-level bootstrap; randomization-inference p-values condition on the original blocked school assignment. Standard errors (S.E.) are reported in parentheses. *, **, and *** denote two-sided conventional p-values below 0.10, 0.05, and 0.01, respectively. Significance is assessed using two-sided, unrounded p-values below 0.10.

Table 5: Application, Admission, and Enrollment Outcomes, by Gender

Outcome	Women: control mean (1)	Women: effect (2)	Men: control mean (3)	Men: effect (4)	Women - men (5)	p-value [q-value] (difference) (6)	Total sample size (7)
Panel A. Overall outcomes across institutions and sources							
Combined application measure	0.294	0.144	1.049	0.912	-0.769	0.393 [0.900]	3,698
		(0.435)		(0.717)	(0.886)		
Administrative applications	0.196	0.246	0.816	0.624	-0.379	0.612 [0.900]	3,698
		(0.354)		(0.611)	(0.739)		
Economics application: broad gateway	0.392	0.538	1.049	0.050	0.488	0.614 [0.900]	3,698
		(0.470)		(0.783)	(0.957)		
Enrollment in an economics degree	0.196	-0.147	0.816	-0.107	-0.040	0.938 [0.974]	3,698
		(0.212)		(0.436)	(0.507)		
Panel B. Universidad Autónoma de Nuevo León (UANL)							
UANL economics exam presented	0.196	0.377	0.466	0.880	-0.503	0.502 [0.900]	3,698
		(0.350)		(0.541)	(0.740)		
UANL economics acceptance	0.098	0.269	0.350	0.453	-0.184	0.760 [0.900]	3,698
		(0.248)		(0.437)	(0.596)		
Panel C. Tecnológico de Monterrey (Tec)							
Enrollment in Tec ECSG gateway	0.196	0.227	0.233	-0.535	0.761	0.125 [0.873]	3,698
		(0.260)		(0.321)	(0.480)		

Notes: Panel A reports outcomes combining institutions or data sources. The combined measure uses the two administrative registries with the endline self-report; the administrative measure combines UANL Economics and Tec LEC applications; broad gateway combines UANL Economics and Tec ECSG applications; and economics-degree enrollment combines Tec economics enrollment with self-reported economics enrollment. This mixed-source enrollment measure records detected events and is not a complete administrative measure. Panel B reports UANL exam participation and acceptance, and Panel C reports Tec ECSG enrollment. Administrative stages are unconditional and equal zero for students not observed at the relevant stage. Estimates come from one pooled treatment-by-gender interaction model per outcome with randomization-block fixed effects and school-clustered standard errors. No specification conditions on August-survey response. Women–men is the interaction coefficient; its conventional q-value retains the Benjamini–Hochberg adjustment across the original family of 14 application and educational-stage outcomes for school assignment. Application outcomes are binary indicators multiplied by 100. Effects and standard errors are reported in percentage points (p.p.); control means are reported as percentages. Total sample size is the sum of the women and men subsamples used in the pooled model. The difference-test column reports the women-minus-men interaction p-value, with its q-value in square brackets. Standard errors (S.E.) are reported in parentheses. *, **, and *** denote two-sided conventional p-values below 0.10, 0.05, and 0.01, respectively. Significance is assessed at 10 percent using unrounded p-values; adjusted tests use q-values below 0.10.

Table 6: Immediate Direct and Spillover Effects, by Gender

Assignment contrast	Men (1)	Women (2)	Women - men (3)	p-value (difference) (4)	Wild bootstrap p-value (5)	Wild bootstrap q-value (6)
Immediate: Beliefs about economics						
Direct treatment vs. control	0.449***	0.716***	0.266***	0.005	0.002	0.010
	(0.090)	(0.087)	(0.087)			
Spillover vs. control	0.084	0.009	-0.074	0.304	0.315	0.420
	(0.065)	(0.066)	(0.071)			
Direct treatment vs. spillover	0.360***	0.714***	0.355***	$<$ 0.001	$<$ 0.001	$<$ 0.001
	(0.067)	(0.085)	(0.091)			
Immediate: Identity and self-concept						
Direct treatment vs. control	0.061	0.274***	0.213**	0.010	0.014	0.027
	(0.072)	(0.047)	(0.077)			
Spillover vs. control	-0.027	0.053	0.080	0.155	0.168	0.337
	(0.044)	(0.037)	(0.055)			
Direct treatment vs. spillover	0.093	0.232***	0.139*	0.081	0.080	0.161
	(0.068)	(0.049)	(0.079)			
Immediate: Intentions and aspirations						
Direct treatment vs. control	-0.042	-0.003	0.039	0.660	0.666	0.724
	(0.116)	(0.097)	(0.088)			
Spillover vs. control	-0.012	-0.065	-0.052	0.505	0.501	0.501
	(0.066)	(0.064)	(0.078)			
Direct treatment vs. spillover	-0.036	0.079	0.116	0.221	0.226	0.301
	(0.070)	(0.084)	(0.094)			
Immediate: Gender stereotypes						
Direct treatment vs. control	0.076	0.039	-0.038	0.717	0.724	0.724
	(0.085)	(0.056)	(0.103)			
Spillover vs. control	0.119**	-0.003	-0.121	0.128	0.152	0.337
	(0.053)	(0.058)	(0.077)			
Direct treatment vs. spillover	-0.030	0.052	0.082	0.472	0.486	0.486
	(0.092)	(0.058)	(0.113)			

Notes: The table reports immediate post-intervention survey outcomes. Men and women effects are estimated jointly in pooled gender-interaction models; Women–men is the interaction coefficient. Direct treatment versus control compares direct-treatment classrooms with classrooms in control schools; spillover versus control compares spillover classrooms in treated schools with classrooms in control schools; and direct treatment versus spillover compares direct-treatment and spillover classrooms within treated schools. The two cross-school contrasts include randomization-block fixed effects and use school-clustered standard errors. The within-school contrast includes school fixed effects and uses classroom-clustered standard errors. Survey-index effects are in standard deviations relative to the baseline control group; positive values indicate more economics-oriented responses or less stereotyped views. The conventional p-value tests the gender interaction using the design-aligned clustered standard error. Wild-bootstrap p-values use a null-imposed wild cluster bootstrap, and Wild-bootstrap q-values control the false-discovery rate within the displayed outcome-by-contrast family. Standard errors (S.E.) are reported in parentheses. *, **, and *** denote two-sided conventional p-values below 0.10, 0.05, and 0.01, respectively. Significance is assessed at 10 percent using unrounded p-values; adjusted tests use q-values below 0.10.

Table 7: Follow-up Direct and Spillover Effects, by Gender

Assignment contrast	Men (1)	Women (2)	Women - men (3)	p-value (difference) (4)	Wild bootstrap p-value (5)	Wild bootstrap q-value (6)
Follow-up: Beliefs about economics						
Direct treatment vs. control	0.172*	0.405***	0.233**	0.039	0.056	0.112
	(0.089)	(0.071)	(0.107)			
Spillover vs. control	0.092	0.019	-0.074	0.442	0.444	0.663
	(0.093)	(0.061)	(0.094)			
Direct treatment vs. spillover	0.156*	0.415***	0.260**	0.025	0.029	0.117
	(0.083)	(0.068)	(0.114)			
Follow-up: Identity and self-concept						
Direct treatment vs. control	-0.087	0.076	0.163**	0.025	0.032	0.112
	(0.063)	(0.064)	(0.068)			
Spillover vs. control	0.045	0.096*	0.050	0.485	0.497	0.663
	(0.055)	(0.056)	(0.071)			
Direct treatment vs. spillover	-0.107	-0.008	0.100	0.253	0.255	0.339
	(0.073)	(0.045)	(0.087)			
Follow-up: Intentions and aspirations						
Direct treatment vs. control	0.073	0.032	-0.041	0.684	0.688	0.863
	(0.084)	(0.056)	(0.099)			
Spillover vs. control	-0.010	0.023	0.033	0.749	0.762	0.762
	(0.083)	(0.057)	(0.103)			
Direct treatment vs. spillover	0.105	0.021	-0.084	0.472	0.477	0.477
	(0.091)	(0.072)	(0.117)			
Follow-up: Gender stereotypes						
Direct treatment vs. control	0.093	0.112	0.019	0.859	0.863	0.863
	(0.090)	(0.076)	(0.107)			
Spillover vs. control	0.144	-0.069	-0.213**	0.039	0.052	0.207
	(0.090)	(0.065)	(0.099)			
Direct treatment vs. spillover	-0.017	0.197***	0.214	0.130	0.131	0.262
	(0.111)	(0.067)	(0.140)			

Notes: The table reports follow-up survey outcomes. Men and women effects are estimated jointly in pooled gender-interaction models; Women–men is the interaction coefficient. Direct treatment versus control compares direct-treatment classrooms with classrooms in control schools; spillover versus control compares spillover classrooms in treated schools with classrooms in control schools; and direct treatment versus spillover compares direct-treatment and spillover classrooms within treated schools. The two cross-school contrasts include randomization-block fixed effects and use school-clustered standard errors. The within-school contrast includes school fixed effects and uses classroom-clustered standard errors. Survey-index effects are in standard deviations relative to the baseline control group; positive values indicate more economics-oriented responses or less stereotyped views. The conventional p-value tests the gender interaction using the design-aligned clustered standard error. Wild-bootstrap p-values use a null-imposed wild cluster bootstrap, and Wild-bootstrap q-values control the false-discovery rate within the displayed outcome-by-contrast family. Standard errors (S.E.) are reported in parentheses. *, **, and *** denote two-sided conventional p-values below 0.10, 0.05, and 0.01, respectively. Significance is assessed at 10 percent using unrounded p-values; adjusted tests use q-values below 0.10.

Table 8: Direct and Spillover Effects on Economics Applications, by Gender

Assignment contrast	Men (1)	Women (2)	Women - men (3)	p-value (difference) (4)	Wild bootstrap p-value (5)	Wild bootstrap q-value (6)
Administrative applications						
Direct treatment vs. control	1.401	0.290	-1.111	0.363	0.505	0.999
	(1.025)	(0.410)	(1.201)			
Spillover vs. control	-0.184	0.223	0.407	0.396	0.409	0.409
	(0.523)	(0.398)	(0.472)			
Direct treatment vs. spillover	1.540	0.052	-1.487	0.150	0.156	0.311
	(0.938)	(0.530)	(1.023)			
Economics application: broad gateway						
Direct treatment vs. control	0.793	0.793	-0.001	1.000	0.999	0.999
	(1.194)	(0.516)	(1.461)			
Spillover vs. control	-0.720	0.329	1.049*	0.075	0.063	0.125
	(0.630)	(0.484)	(0.567)			
Direct treatment vs. spillover	1.509	0.367	-1.142	0.399	0.422	0.422
	(1.010)	(0.744)	(1.348)			

Notes: Men and women effects are estimated jointly in pooled gender-interaction models; Women–men is the interaction coefficient. Direct treatment versus control compares direct-treatment classrooms with classrooms in control schools; spillover versus control compares spillover classrooms in treated schools with classrooms in control schools; and direct treatment versus spillover compares direct-treatment and spillover classrooms within treated schools. The two cross-school contrasts include randomization-block fixed effects and use school-clustered standard errors. The within-school contrast includes school fixed effects and uses classroom-clustered standard errors. Application outcomes are binary indicators multiplied by 100. Effects and standard errors are reported in percentage points (p.p.). The administrative measure equals one for an application to UANL Economics or Tec LEC. The conventional p-value tests the gender interaction using the design-aligned clustered standard error. Wild-bootstrap p-values use a null-imposed wild cluster bootstrap, and Wild-bootstrap q-values control the false-discovery rate within the displayed application-by-contrast family. Because the administrative outcome is rare, both point estimates and confidence intervals should be interpreted cautiously. Standard errors (S.E.) are reported in parentheses. *, **, and *** denote two-sided conventional p-values below 0.10, 0.05, and 0.01, respectively. Significance is assessed at 10 percent using unrounded p-values; adjusted tests use q-values below 0.10.

Table 9: Mechanism-Relevant Heterogeneity by Baseline Beliefs and Confidence

Outcome	Low: control mean (1)	Low: effect (2)	High: control mean (3)	High: effect (4)	High - low (5)	p-value [wild q-value] (difference) (6)	Sample size (7)
Panel A. Baseline economics-beliefs index: Low beliefs versus High beliefs							
Immediate: Beliefs about economics	-0.205	0.376***	0.387	0.227***	-0.149*	0.053 [0.396]	3,350
		(0.069)		(0.071)	(0.073)		
Follow-up: Beliefs about economics	-0.285	0.168**	0.120	0.178***	0.011	0.892 [0.977]	1,850
		(0.076)		(0.050)	(0.079)		
Administrative applications	0.000	0.200	0.417	0.708	0.509	0.306 [0.401]	1,949
		(0.157)		(0.517)	(0.487)		
Panel B. Baseline interest in economics: Low interest versus High interest							
Immediate: Beliefs about economics	-0.114	0.445***	0.223	0.221***	-0.223***	0.008 [0.110]	3,350
		(0.080)		(0.064)	(0.078)		
Follow-up: Beliefs about economics	-0.176	0.167**	-0.017	0.169***	0.002	0.983 [0.982]	1,850
		(0.078)		(0.056)	(0.085)		
Administrative applications	0.000	0.012	0.375	0.736	0.724	0.162 [0.400]	2,009
		(0.118)		(0.489)	(0.504)		
Panel C. Baseline self-assessed mathematics ability: Low math confidence versus High math confidence							
Immediate: Beliefs about economics	0.019	0.288***	0.179	0.309***	0.021	0.781 [0.978]	3,349
		(0.064)		(0.074)	(0.076)		
Follow-up: Beliefs about economics	-0.147	0.152**	-0.015	0.178***	0.026	0.716 [0.977]	1,849
		(0.065)		(0.058)	(0.070)		
Administrative applications	0.000	0.039	0.463	0.812	0.773	0.178 [0.400]	2,008
		(0.098)		(0.569)	(0.560)		

Notes: Panels A–C define low and high subgroups using the baseline economics-beliefs index, baseline interest in economics, and baseline self-assessed mathematics ability, respectively. Each row is estimated from a separate pooled model containing treated-school assignment, the displayed moderator, and their interaction. Low and high control means are unadjusted subgroup means; the subgroup effects are linear combinations from the pooled model; and High–low is the formal interaction estimate. Models include randomization-block fixed effects and use school-clustered standard errors. Survey-index effects and control means are in standard deviations relative to the baseline control group; positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100. Effects and standard errors are reported in percentage points (p.p.); control means are reported as percentages. The wild-bootstrap q-value controls the false-discovery rate within the prespecified interaction family. These baseline moderators are mechanism-relevant sources of heterogeneity, not mediators, and the estimates do not identify causal pathways. The difference-test column reports the conventional interaction p-value and the wild-bootstrap q-value in square brackets. Standard errors (S.E.) are reported in parentheses. *, **, and *** denote two-sided conventional p-values below 0.10, 0.05, and 0.01, respectively. Significance is assessed at 10 percent using unrounded p-values; adjusted tests use q-values below 0.10.

Table 10: Heterogeneity by Predetermined Student Characteristics

Outcome	Low: control mean (1)	Low: effect (2)	High: control mean (3)	High: effect (4)	High - low (5)	p-value [wild q-value] (difference) (6)	Sample size (7)
Panel A. Student gender: Male versus Female							
Immediate: Beliefs about economics	0.113	0.268***	0.094	0.342***	0.074	0.289 [0.477]	3,306
		(0.071)		(0.067)	(0.069)		
Follow-up: Beliefs about economics	-0.083	0.127*	-0.063	0.194***	0.067	0.362 [0.500]	1,823
		(0.074)		(0.053)	(0.072)		
Administrative applications	0.816	0.624	0.196	0.246	-0.379	0.612 [0.696]	3,698
		(0.611)		(0.354)	(0.739)		
Panel B. At least one university-educated parent: No university-educated parent versus University-educated parent							
Immediate: Beliefs about economics	0.118	0.311***	0.102	0.284***	-0.027	0.723 [0.869]	3,278
		(0.072)		(0.064)	(0.076)		
Follow-up: Beliefs about economics	-0.046	0.164***	-0.097	0.157**	-0.007	0.935 [0.995]	1,811
		(0.057)		(0.075)	(0.084)		
Administrative applications	0.187	0.228	0.217	0.675	0.447	0.554 [0.950]	1,974
		(0.413)		(0.513)	(0.746)		
Panel C. School mean parental university education: Low school parental education versus High school parental education							
Immediate: Beliefs about economics	0.089	0.320***	0.110	0.288***	-0.032	0.802 [0.869]	3,350
		(0.091)		(0.079)	(0.127)		
Follow-up: Beliefs about economics	-0.064	0.108	-0.087	0.259***	0.152	0.161 [0.953]	1,850
		(0.071)		(0.076)	(0.105)		
Administrative applications	0.263	0.661	0.156	-0.011	-0.671	0.460 [0.950]	2,009
		(0.569)		(0.476)	(0.895)		
Panel D. PAP household-asset SES index: Low SES assets versus High SES assets							
Immediate: Beliefs about economics	0.047	0.279**	0.127	0.313***	0.034	0.701 [0.869]	3,278
		(0.102)		(0.053)	(0.087)		
Follow-up: Beliefs about economics	-0.174	0.172**	-0.036	0.169**	-0.003	0.972 [0.995]	1,806
		(0.064)		(0.069)	(0.087)		
Administrative applications	0.000	0.444	0.285	0.396	-0.048	0.934 [0.950]	1,976
		(0.482)		(0.328)	(0.577)		
Panel E. Self-reported baseline GPA: Low GPA versus High GPA							
Immediate: Beliefs about economics	0.083	0.244***	0.125	0.362***	0.118*	0.090 [0.288]	3,331
		(0.073)		(0.062)	(0.067)		
Follow-up: Beliefs about economics	-0.113	0.175***	-0.037	0.159**	-0.015	0.866 [0.867]	1,841
		(0.059)		(0.076)	(0.089)		
Administrative applications	0.000	0.520*	0.402	0.274	-0.246	0.604 [0.630]	1,998
		(0.272)		(0.423)	(0.469)		

Notes: The panels report heterogeneity by predetermined student characteristics. Each row is estimated from a separate pooled model containing treated-school assignment, the displayed moderator, and their interaction. Low and high control means are unadjusted subgroup means; subgroup effects are linear combinations from the pooled model; and High–low is the formal interaction estimate. All models include randomization-block fixed effects and use school-clustered standard errors. Survey-index effects and control means are in standard deviations relative to the baseline control group; positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100. Effects and standard errors are reported in percentage points (p.p.); control means are reported as percentages. The wild-bootstrap q-value controls the false-discovery rate within the prespecified interaction family. The table displays immediate beliefs, follow-up beliefs, and the administrative application measure; complete outcome-by-moderator families appear in the Online Appendix. The difference-test column reports the conventional interaction p-value and the wild-bootstrap q-value in square brackets. Standard errors (S.E.) are reported in parentheses. *, **, and *** denote two-sided conventional p-values below 0.10, 0.05, and 0.01, respectively. Significance is assessed at 10 percent using unrounded p-values; adjusted tests use q-values below 0.10.

Table 11: Heterogeneity by Classroom and Intervention Context

Outcome	Low: control mean (1)	Low: effect (2)	High: control mean (3)	High: effect (4)	High - low (5)	p-value [wild q-value] (difference) (6)
Panel A. Baseline classroom female share: Low female share versus High female share						
Immediate: Beliefs about economics	0.088	0.423***	0.116	0.194**	-0.228**	0.023 [0.187]
		(0.076)		(0.072)	(0.094)	
Follow-up: Beliefs about economics	-0.072	0.221**	-0.085	0.130*	-0.091	0.377 [0.716]
		(0.082)		(0.066)	(0.101)	
Administrative applications	0.264	0.964	0.156	0.160	-0.804	0.291 [0.762]
		(0.617)		(0.318)	(0.746)	
Panel B. Baseline classroom mean GPA: Lower two classroom terciles versus Top classroom GPA tercile						
Immediate: Beliefs about economics	0.119	0.287***	0.049	0.346***	0.059	0.542 [0.784]
		(0.071)		(0.078)	(0.096)	
Follow-up: Beliefs about economics	-0.062	0.156**	-0.145	0.218***	0.062	0.451 [0.724]
		(0.060)		(0.072)	(0.081)	
Administrative applications	0.129	0.368	0.405	0.479	0.111	0.895 [0.903]
		(0.334)		(0.697)	(0.837)	
Panel C. Baseline classroom men-better stereotype: Low stereotypes versus High stereotypes						
Immediate: Beliefs about economics	0.138	0.367***	0.069	0.203**	-0.164	0.104 [0.362]
		(0.072)		(0.085)	(0.098)	
Follow-up: Beliefs about economics	-0.081	0.221***	-0.076	0.086	-0.135	0.185 [0.545]
		(0.078)		(0.071)	(0.099)	
Administrative applications	0.180	0.303	0.216	0.549	0.246	0.755 [0.903]
		(0.388)		(0.570)	(0.779)	
Panel D. Female classroom teacher: Male teacher versus Female teacher						
Immediate: Beliefs about economics	0.061	0.093	0.103	0.438***	0.345***	$<$ 0.001 [0.112]
		(0.065)		(0.060)	(0.092)	
Follow-up: Beliefs about economics	-0.174	0.192**	-0.059	0.193***	0.001	0.992 [0.992]
		(0.092)		(0.057)	(0.086)	
Administrative applications	0.000	2.083**	0.254	-0.174	-2.257**	0.020 [0.418]
		(0.887)		(0.230)	(0.914)	

Notes: The panels define the classroom context using the baseline female share, classroom mean GPA, and classroom endorsement of the stereotype that men are naturally better at mathematics. The subgroup labels and cutoffs shown in each panel follow the prespecified construction. Each row comes from a pooled treated-school-assignment interaction model. Low and high control means are unadjusted subgroup means; subgroup effects are linear combinations from that model; and High–low is the interaction coefficient. Models include randomization-block fixed effects and use school-clustered standard errors. Survey-index effects and control means are in standard deviations relative to the baseline control group; positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100. Effects and standard errors are reported in percentage points (p.p.); control means are reported as percentages. The wild-bootstrap q-value controls the false-discovery rate within the prespecified classroom-moderator family. Because the number of schools supporting some assignment-by-subgroup cells is small, the interaction estimates should be interpreted cautiously. The difference-test column reports the conventional interaction p-value and the wild-bootstrap q-value in square brackets. Standard errors (S.E.) are reported in parentheses. *, **, and *** denote two-sided conventional p-values below 0.10, 0.05, and 0.01, respectively. Significance is assessed at 10 percent using unrounded p-values; adjusted tests use q-values below 0.10.