**Supplementary Analyses to Accompany**

# "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

*Not for Publication*

# Contents

**Implementation of the Pre-Analysis Plan** **S-6**

**Supplementary Figures** **S-8**

**A. Survey Outcomes and Persistence** **S-19**

**B. Application, Admission, and Enrollment Outcomes** **S-34**

**C. Implementation, Information, and Diffusion** **S-40**

**D. Academic Expectations, Performance, and Self-Assessment** **S-54**

**E. Exploratory Mediation** **S-60**

**F. Heterogeneity by Student Characteristics and Classroom Context** **S-64**

**G. Heterogeneity by Social, Time, and Risk Preferences** **S-76**

**H. Machine-Learning Heterogeneity** **S-84**

**I. Response, Attrition, and Placebo Tests** **S-92**

**J. Additional Inference Comparisons** **S-99**

# List of Figures


| | | |
|---|---|---|
| S1 | Classroom-Level Baseline Participation | S-8 |
| S2 | Gender Differences in Treatment Effects: Immediate and Follow-Up | S-9 |
| S3 | Applications to Economics by Experimental Arm | S-10 |
| S4 | Applications to Economics by Student Gender | S-11 |
| S5 | Survey-Index Treatment Effects by Baseline Grade Point Average | S-12 |
| S6 | Economics-Application Treatment Effects by Baseline Grade Point Average | S-13 |
| S7 | Treatment-Effect Differences in Beliefs by Classroom Context | S-14 |
| S8 | Treatment-Effect Differences in Economics Applications by Classroom Context | S-15 |
| S9 | Treatment-Effect Differences in Beliefs by Classroom Teacher Gender | S-16 |
| S10 | Economics-Application Effect Differences by Classroom Teacher Gender | S-17 |
| S11 | Randomization-Inference Null Distributions | S-18 |

# List of Tables


| | | |
|---|---|---|
| | Planned and Implemented Analyses | S-6 |
| S1 | Immediate PAP Item Effects by Gender | S-19 |
| S2 | Follow-Up PAP Item Effects by Gender: Beliefs about Economics | S-21 |
| S3 | Follow-Up PAP Item Effects by Gender: Identity and Self-Concept | S-23 |
| S4 | Follow-Up PAP Item Effects by Gender: Intentions and Stereotypes | S-25 |
| S5 | Gender Differences in Immediate Effects | S-27 |
| S6 | Gender Differences in Follow-Up Effects | S-29 |
| S7 | Gender Differences in Persistence: Follow-Up Minus Immediate | S-31 |
| S8 | Economics Enrollment and UANL Acceptance: School-Assignment Contrasts | S-34 |
| S9 | Economics Enrollment and UANL Acceptance: Within-Treated-School Contrast | S-35 |
| S10 | Application, Admission, and Enrollment Outcomes by Gender | S-36 |
| S11 | Predetermined Covariates of Economics Applications | S-38 |
| S12 | Reported Attendance and Reception of the Visit by Gender | S-40 |
| S13 | Recall, Peer Discussion, and Family Diffusion | S-41 |
| S14 | Exposure and Recall by Gender | S-43 |
| S15 | Diffusion and Follow-Up Information Requests by Gender | S-45 |
| S16 | Information and Perceived-Barriers Mechanisms | S-47 |
| S17 | Perceived Barriers by Gender | S-49 |
| S18 | Information Requests by Gender | S-52 |
| S19 | Treatment Effects on Academic Expectations and Quiz Performance | S-54 |
| S20 | Treatment Effects on Academic Expectations by Gender | S-55 |
| S21 | Realized Quantitative Performance by Gender | S-56 |
| S22 | Treatment Effects on Post-Intervention Mathematics Self-Assessment | S-58 |
| S23 | Treatment Effects on Mathematics Self-Assessment by Gender | S-59 |
| S24 | Exploratory Survey Mediation with Beliefs about Economics as Mediator | S-60 |
| S25 | Exploratory Survey Mediation with Identity and Self-Concept as Mediator | S-61 |
| S26 | Exploratory Survey Mediation with Intentions and Aspirations as Mediator | S-62 |
| S27 | Exploratory Survey Mediation with Gender Stereotypes as Mediator | S-63 |

| | | |
|---|---|---|
| S28 | Heterogeneity by Parental Education | S-64 |
| S29 | Heterogeneity by School-Level Parental Education | S-65 |
| S30 | Heterogeneity by Household Socioeconomic Status | S-66 |
| S31 | Heterogeneity by Baseline Academic Performance | S-67 |
| S32 | Heterogeneity by Baseline Classroom Female Share | S-68 |
| S33 | Heterogeneity by Baseline Classroom Mean GPA | S-69 |
| S34 | Heterogeneity by Baseline Classroom Gender Stereotypes | S-70 |
| S35 | Exploratory Heterogeneity by Classroom Teacher Gender | S-71 |
| S36 | Gendered Heterogeneity by Baseline Beliefs and Interest | S-72 |
| S37 | Gendered Heterogeneity by Baseline Mathematics Self-Assessment and Teacher Gender | S-74 |
| S38 | Heterogeneity by Altruism | S-76 |
| S39 | Heterogeneity by Quantitative Altruism | S-77 |
| S40 | Heterogeneity by Generalized Trust | S-78 |
| S41 | Heterogeneity by Willingness to Wait | S-79 |
| S42 | Heterogeneity by Patience Valuation | S-80 |
| S43 | Heterogeneity by Procrastination | S-81 |
| S44 | Heterogeneity by Willingness to Take Risks | S-82 |
| S45 | Heterogeneity by Risk Valuation | S-83 |
| S46 | Machine-Learning Heterogeneity in Follow-up Survey Outcomes | S-84 |
| S47 | Machine-Learning Heterogeneity in Economics Applications | S-85 |
| S48 | Machine-Learning Heterogeneity by Student Gender | S-86 |
| S49 | Predetermined Characteristics Associated with Model-Implied CATE Scores: Beliefs about Economics | S-87 |
| S50 | Predetermined Characteristics Associated with Model-Implied CATE Scores: Identity and Self-Concept | S-88 |
| S51 | Predetermined Characteristics Associated with Model-Implied CATE Scores: Intentions and Aspirations | S-89 |
| S52 | Predetermined Characteristics Associated with Model-Implied CATE Scores: Gender Stereotypes | S-90 |

S53 Predetermined Characteristics Associated with Model-Implied CATE Scores: Economics Applications . . . . . S-91
S54 Survey Response by Gender . . . . . S-92
S55 Gender-Specific Estimates with Inverse-Probability Weighting . . . . . S-94
S56 Placebo Outcomes and Subject-Specific Benchmarks by Gender . . . . . S-96
S57 School-ITT Inference under Alternative Clustering . . . . . S-99

## Implementation of the Pre-Analysis Plan

The pre-analysis plan of October 20, 2025 specified the main outcomes and analyses of mechanisms, heterogeneity, and robustness. We report the planned analyses alongside the specifications we implemented. Where the data or field procedures required a change, we explain it below. Several analyses also include subgroup comparisons or specifications developed after registration.

Planned and Implemented Analyses

| **Planned analysis** | **Implementation** | **Where reported** |
|---|---|---|
| *Panel A. Study design and outcome measurement* | | |
| Assignment and study population<br>*Plan § 8* | We randomized schools and then classrooms within treated schools. The eligible population includes students in the final semester of four- and six-semester programs. | Main text, Tables 1–2<br>Figure A5, Table A2 |
| Economics applications and cohorts<br>*Plan §§ 9.1, 11.1* | We combine university records and August survey reports and also report each source separately. Differences in consent across the 2025 and 2026 cohorts preclude the planned difference-in-differences comparison. The UANL records identify applications and acceptance but do not establish enrollment. | Main text, Section 7.3<br>Tables A15–A16 |
| Survey outcomes and index construction<br>*Plan §§ 9.2, 10.3–10.5* | We report effects for the four survey domains and document the items available at each wave, their reliability, and the construction of each index. The appendix reproduces all four survey instruments. | Main text, Table 3<br>Appendix sections B and D<br>Tables A5–A7, A8–A11 |
| *Panel B. Treatment effects and heterogeneity* | | |
| Survey adjustment and spillovers<br>*Plan §§ 11.2, 13.6* | We adjust survey estimates for baseline outcomes and distinguish assignment to a treated school from direct classroom exposure and spillovers. We also vary the covariate adjustment. | Main text, Tables 6–8<br>Tables A18–A20, A30–A33 |
| Heterogeneity by baseline characteristics<br>*Plan §§ 4.1, 5, 12.1* | We compare effects by gender, initial beliefs and interest, mathematics confidence, socioeconomic background, grades, and classroom composition. Additional analyses examine teacher gender and how subgroup differences vary by student gender. | Main text, Section 9<br>Tables A21–A24<br>Tables S28–S34, S35–S37 |
| Machine-learning analysis of heterogeneity<br>*Plan § 12.2* | We estimate treatment-effect heterogeneity using LASSO and forests. LASSO folds are defined by school and forest folds by observation. Calibration tests inform the interpretation of the predicted subgroup profiles. | Tables S46–S53 |

*Planned and Implemented Analyses (continued)*

| Planned analysis | Implementation | Where reported |
|---|---|---|
| *Panel C. Mechanisms* | | |
| Mediation<br>*Plan § 14.2* | We estimate a separate indirect effect for each survey domain. A causal interpretation requires sequential ignorability of the mediator–outcome relationship. We report these estimates as exploratory and interpret each separately. | Main text, Section 8.2<br>Tables S24–S27 |
| Laboratory and behavioral measures<br>*Plan § 15* | The surveys measured academic expectations, quiz performance, self-assessments, and preferences. We did not implement the planned laboratory manipulations of information and identity. The expectations and quiz questions use different benchmarks, which precludes estimating a calibration gap. | Tables S19–S23, S38–S45 |
| *Panel D. Robustness and inference* | | |
| Attrition and sample selection<br>*Plan § 13.1* | We compare response rates across arms and assess sensitivity using inverse-probability weighting, Lee bounds, and a balanced panel. The supplement extends the response and weighting analyses by gender. | Main text, Section 10.1<br>Tables A25–A29<br>Tables S54–S55 |
| Alternative outcomes and specifications<br>*Plan §§ 13.2, 13.6* | We vary the covariates, index construction, and coverage of economics programs and data sources. We report admission and enrollment outcomes separately, along with descriptive associations between applications and baseline characteristics. | Main text, Section 10.2<br>Tables A30–A37<br>Tables S8–S10, S11 |
| Balance, placebos, and inference<br>*Plan §§ 13.3–13.5* | We report baseline balance and distinguish placebo subjects from the mathematics and economics benchmarks. We compare inference methods, alternative levels of clustering, and placebo and benchmark effects by gender. | Main text, Table 1 and Section 10.3<br>Tables A38–A39<br>Tables S56–S57 |

*Notes:* Plan section numbers refer to the pre-analysis plan. The A prefix denotes the Online Appendix and the S prefix denotes this supplement. Additional item-level comparisons by gender and tests of persistence appear in Tables S1–S7. Tables S12–S18 cover implementation, recall, diffusion, and perceived barriers. These extensions retain the original testing families.

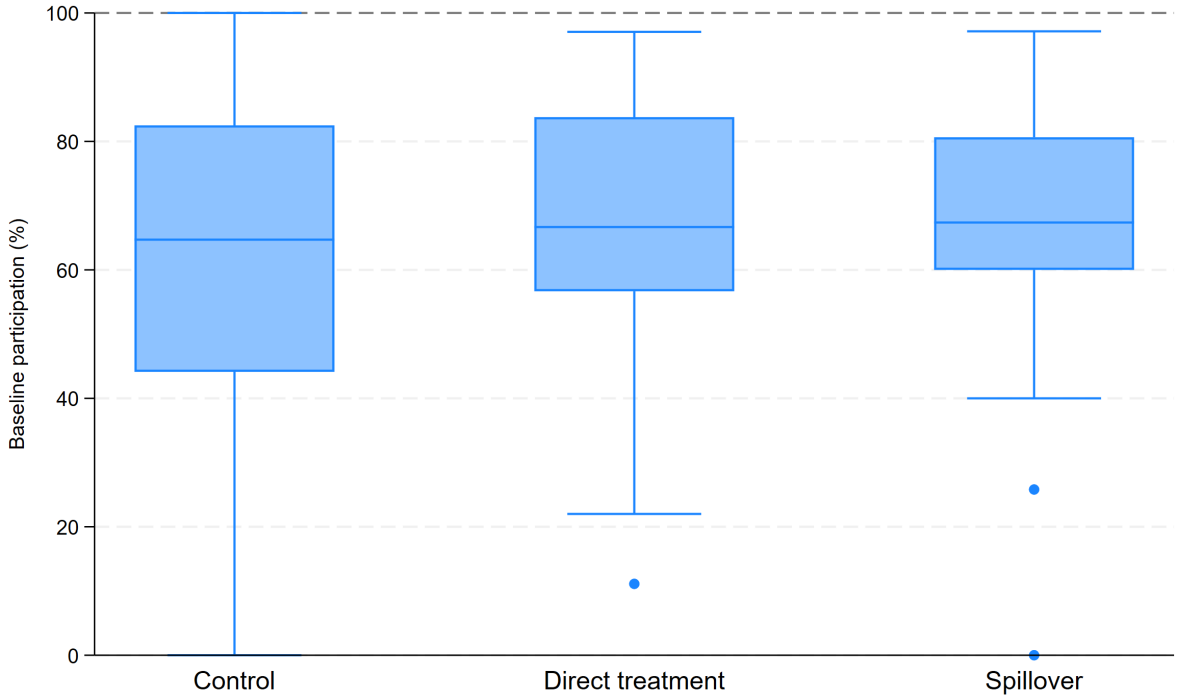



Figure S1: Classroom-Level Baseline Participation

*Notes:* The figure shows baseline participation across all 168 assigned classrooms, including the nine with no completed baseline surveys. For each classroom, we divide the number of completed surveys by enrollment on the assignment roster and express the result as a percentage. Boxes show the interquartile range, the line inside each box marks the median, and dots mark observations beyond the whiskers. In one control classroom, 23 students completed the survey although the roster listed 21. We display the upper whisker at 100% and retain the original counts in the data.

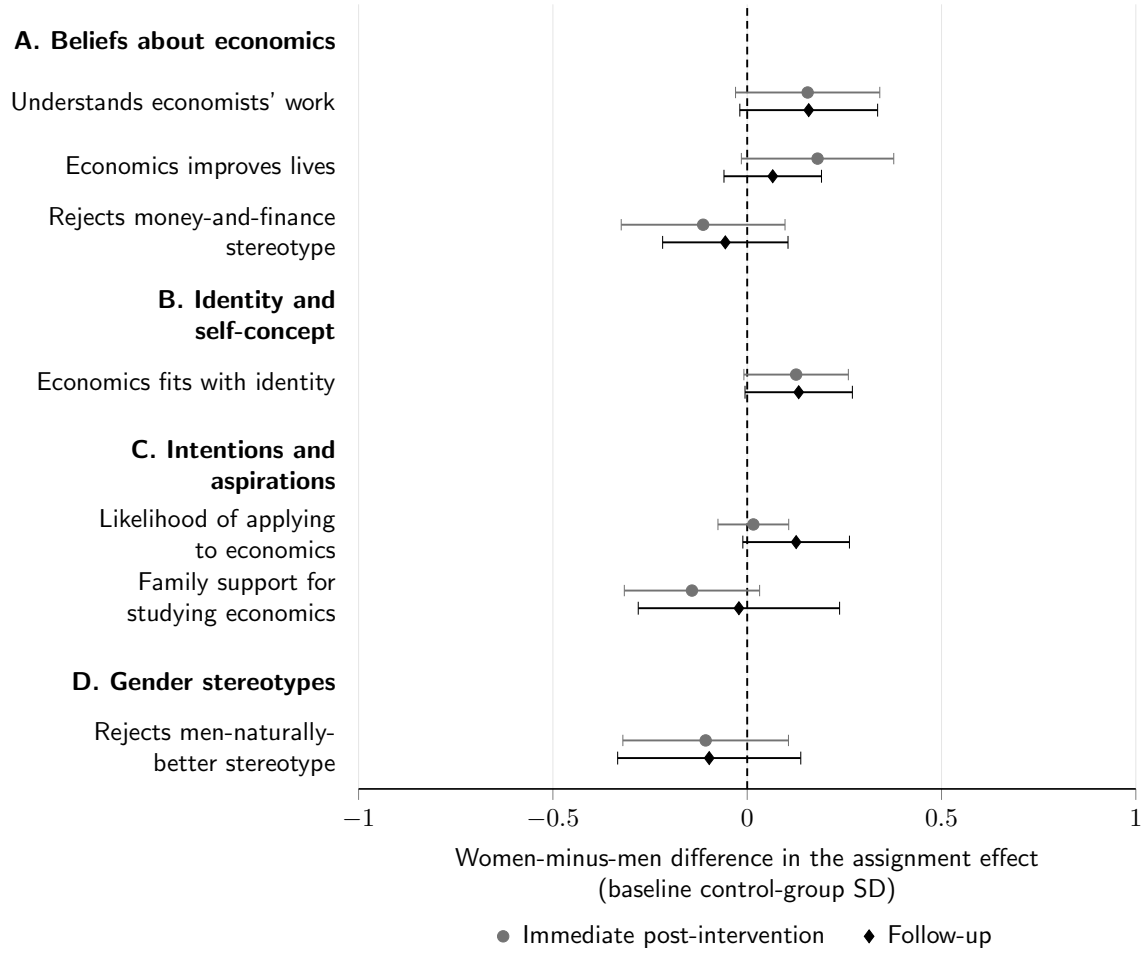



Figure S2: Gender Differences in Treatment Effects: Immediate and Follow-Up

*Notes:* The sample comprises final-year students from four- and six-semester programs with baseline, immediate post-intervention, and follow-up responses to the corresponding item and observed baseline gender. The same students contribute to both waves within each item; samples may differ across items. Panels group the seven common items into beliefs about economics, identity and self-concept, intentions and aspirations, and gender stereotypes. Each point is the women-minus-men difference in the treated-school assignment effect from a pooled model with assignment, gender, and their interaction, estimated separately by item and wave. Treated-school assignment combines directly treated and spillover classrooms and compares them with control schools. Models include randomization-block fixed effects and the item's baseline value, with standard errors clustered at the school assignment unit. Outcomes at both waves use the corresponding baseline control-group mean and standard deviation; the two stereotype-rejection items are reverse-coded. Circles denote immediate effects and diamonds follow-up effects. Horizontal bars are pointwise 95 percent confidence intervals computed as the coefficient plus or minus the school-clustered standard error times the Student-$t$ critical value, using the number of school clusters minus one degrees of freedom (27 for each estimate). These intervals are not adjusted for multiple testing. The dashed line denotes equal effects for women and men; positive values indicate a larger treatment effect for women, not necessarily a positive effect for either gender. The common scale runs from $-1$ to 1 baseline control-group standard deviations. Comparing the two waves visually does not test whether their gender differentials changed: Table S7 reports the formal follow-up-minus-immediate contrast. Follow-up timing differs across treatment arms, and the joint gender-by-persistence analysis was not registered in the PAP. Source: survey responses and the treated-school assignment estimates in Tables S6 and S5.

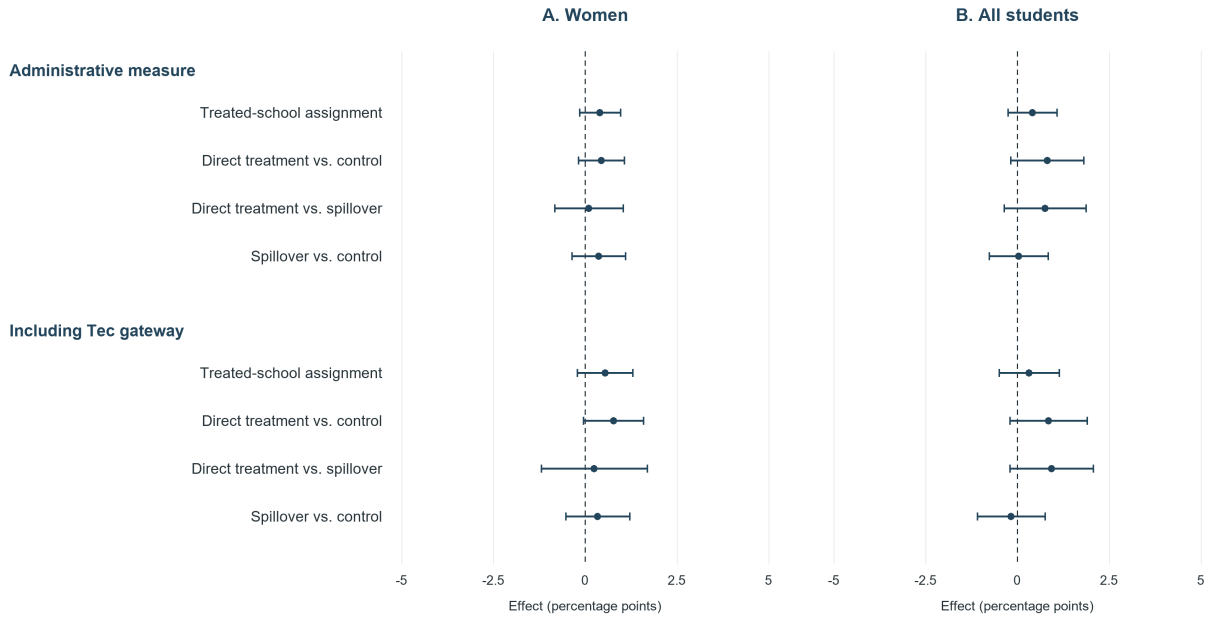



Figure S3: Applications to Economics by Experimental Arm

*Notes:* The left panel uses women in the final year of four- and six-semester programs; the right panel uses all students in that final-year population. Rows distinguish the administrative application measure—an application to UANL Economics or Tec LEC—from the broad gateway definition that additionally counts Tec ECSG. Markers report assignment-aligned differences and horizontal bars 95 percent confidence intervals; the dashed vertical line denotes zero. Treated-school assignment combines direct and spillover classrooms in treated schools. Direct treatment versus control and spillover versus control include the original randomization-block fixed effects and use school-clustered standard errors. Direct treatment versus spillover is identified within treated schools with school fixed effects and classroom-clustered standard errors. The binary outcomes are multiplied by 100, so effects are in percentage points. Both panels use the common $-5$ to 5 scale.

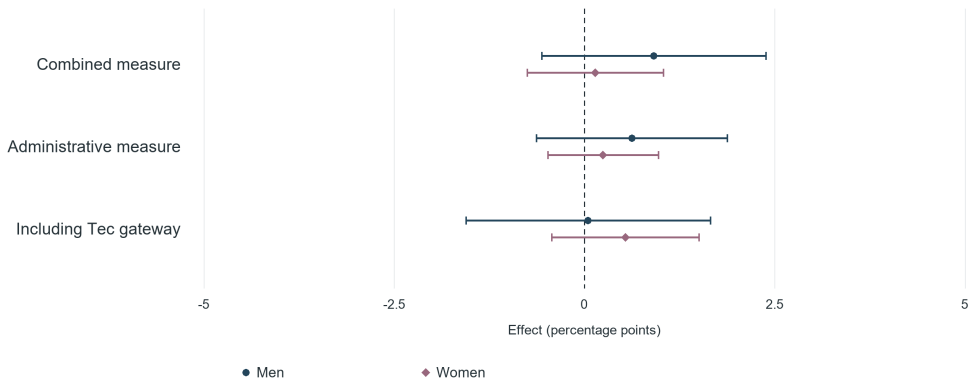



Figure S4: Applications to Economics by Student Gender

*Notes:* The combined measure uses registry applications with economics-degree reports from the August online survey. It records detected events in the full eligible sample; no specification conditions on survey response. A zero indicates no detected event, and incomplete survey coverage can affect event detection. The other rows distinguish the administrative economics-application outcome—an application to UANL Economics or Tec LEC—from the broad gateway definition that additionally counts Tec ECSG. Markers report treated-school assignment effects for men and women as linear combinations from a single pooled treatment-by-gender interaction model; horizontal bars are 95 percent confidence intervals and the dashed vertical line denotes zero. Each specification includes the original randomization-block fixed effects and uses school-clustered standard errors. The binary outcomes are multiplied by 100, so effects are in percentage points. The horizontal scale runs from $-5$ to 5 percentage points; confidence intervals extending beyond this display range are truncated at the boundary. Because positive application events are rare, especially in the administrative measure, both subgroup estimates and gender differences should be interpreted cautiously.

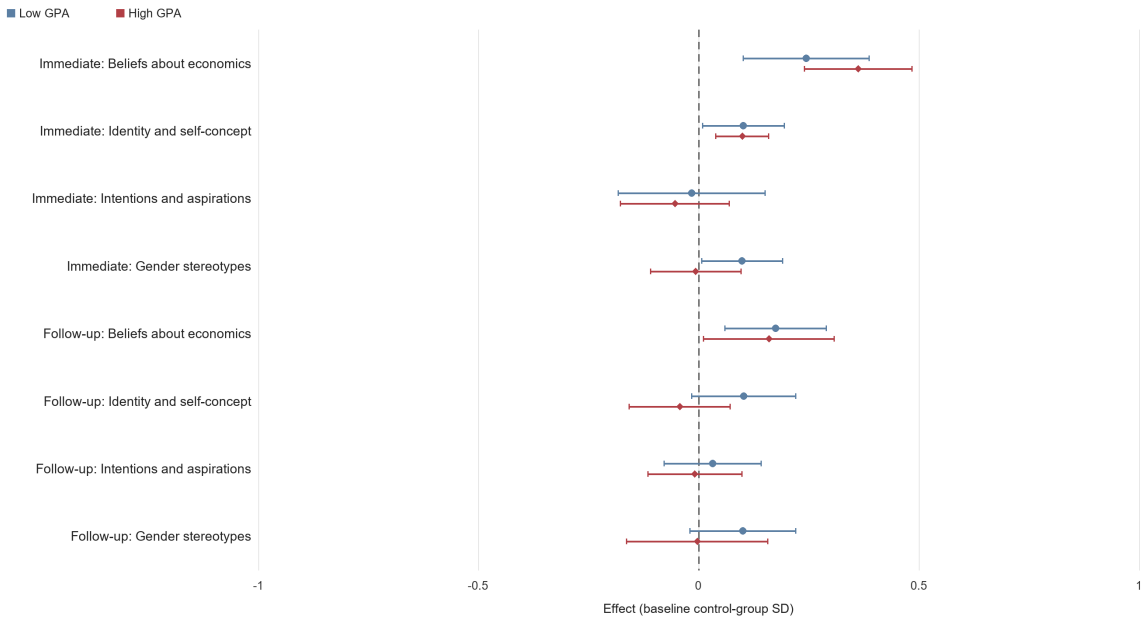



Figure S5: Survey-Index Treatment Effects by Baseline Grade Point Average

*Notes:* The figure reports heterogeneity by predetermined baseline grade point average. For each immediate and follow-up survey-index family, separate markers show the treated-school assignment effect below and above the preregistered GPA cutoff; horizontal bars are 95 percent confidence intervals and the dashed vertical line denotes zero. Subgroup effects are linear combinations from a pooled model containing assignment, the high-GPA indicator, and their interaction. Every model includes the original randomization-block fixed effects and uses school-clustered standard errors. Outcomes are standardized using baseline control-group moments and oriented so that positive values indicate more economics-oriented responses or less stereotyped views. The common survey-index scale matches the other survey-effect figures. The distance between subgroup markers is descriptive; formal high-minus-low interaction tests and multiplicity-adjusted values appear in the corresponding table.

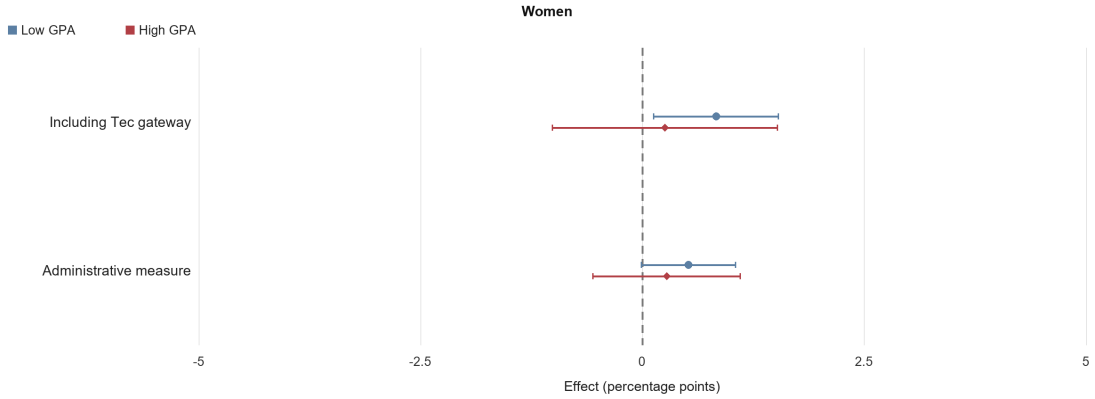



Figure S6: Economics-Application Treatment Effects by Baseline Grade Point Average

*Notes:* The sample consists of women in the final semester, split below and above the preregistered baseline-GPA cutoff. Rows distinguish the administrative application measure—an application to UANL Economics or Tec LEC—from the broad gateway definition that additionally counts Tec ECSG. Subgroup markers are linear combinations from a pooled treated-school-assignment interaction model; horizontal bars are 95 percent confidence intervals and the dashed vertical line denotes zero. Models include the original randomization-block fixed effects and use school-clustered standard errors. The binary outcomes are multiplied by 100, so effects are in percentage points, and the application scale matches the other administrative-outcome figures. Formal high-minus-low interaction tests are reported in the corresponding table. Sparse application events make the subgroup confidence intervals wide.

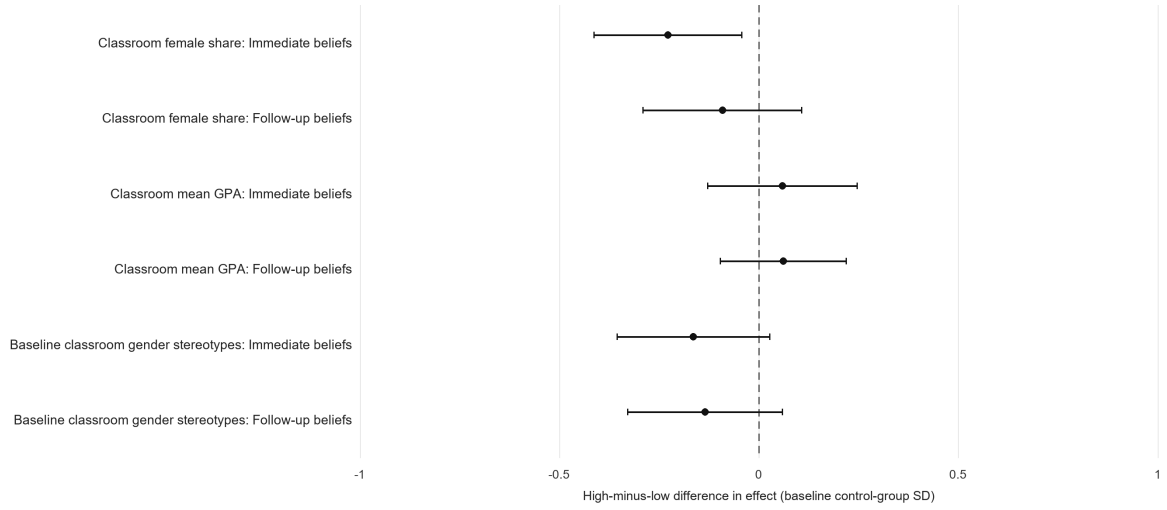



Figure S7: Treatment-Effect Differences in Beliefs by Classroom Context

*Notes:* Rows report heterogeneity in the beliefs-about-economics index by three predetermined classroom characteristics: baseline female share, mean GPA, and endorsement of the stereotype that men are naturally better at mathematics. Each marker is the high-minus-low difference in the treated-school assignment effect from a pooled interaction model; a positive value means a larger effect in the high subgroup. Horizontal bars are 95 percent confidence intervals and the dashed vertical line denotes equal effects. Immediate and follow-up outcomes are displayed separately. Models include the original randomization-block fixed effects and use school-clustered standard errors. Beliefs are standardized using baseline control-group moments and oriented so that positive values indicate more economics-oriented responses. The common survey-index scale matches other survey figures. A flag identifies moderators with fewer than three school clusters in at least one assignment-by-subgroup cell, for which conventional confidence intervals are especially fragile.

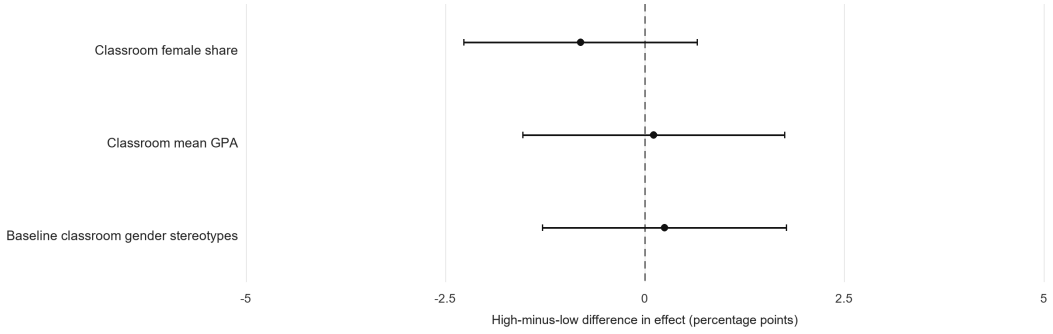



Figure S8: Treatment-Effect Differences in Economics Applications by Classroom Context

*Notes:* The sample consists of women in the final semester. Rows report heterogeneity in the administrative economics-application outcome by three predetermined classroom characteristics: baseline female share, mean GPA, and endorsement of the stereotype that men are naturally better at mathematics. Each marker is the high-minus-low difference in the treated-school assignment effect from a pooled interaction model; a positive value means a larger effect in the high subgroup. Horizontal bars are 95 percent confidence intervals and the dashed vertical line denotes equal effects. Models include the original randomization-block fixed effects and use school-clustered standard errors. The binary application indicator is multiplied by 100, so differences are in percentage points, on the same scale used in the other application figures. A flag identifies moderators with fewer than three school clusters in at least one assignment-by-subgroup cell. Sparse events and limited cluster support warrant cautious interpretation.

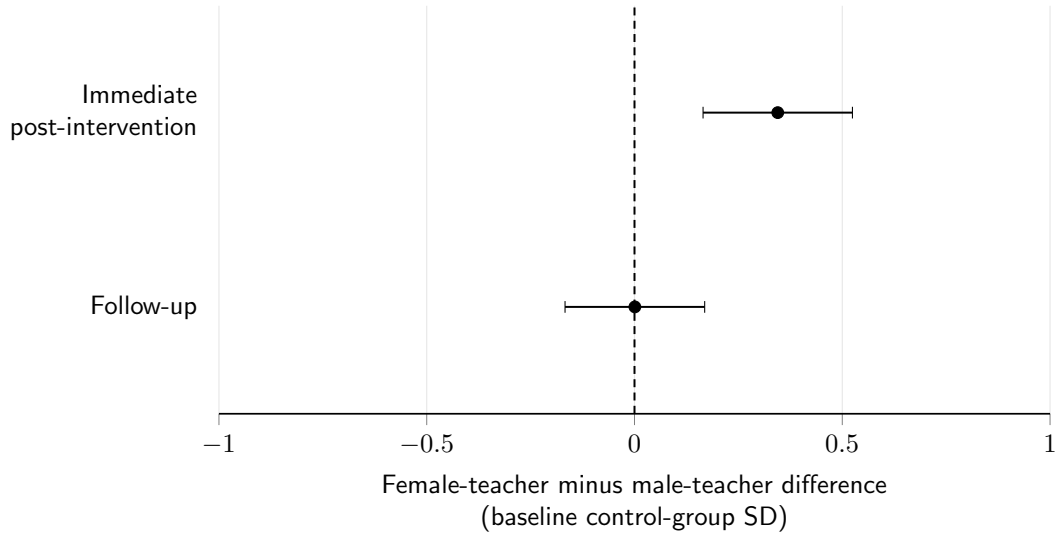



Figure S9: Treatment-Effect Differences in Beliefs by Classroom Teacher Gender

*Notes:* The sample consists of final-year students from four- and six-semester programs observed at the corresponding survey wave. Rows distinguish immediate post-intervention and follow-up effects on the beliefs-about-economics index, standardized using baseline control-group moments. Each point is the female-teacher-minus-male-teacher difference in the treated-school assignment effect from a pooled model with assignment, teacher gender, and their interaction. Treated-school assignment combines directly treated and spillover classrooms and compares them with control schools. Models include the original randomization-block fixed effects and the index's baseline value, with standard errors clustered at the school assignment unit. Horizontal bars are pointwise, normal-approximation 95 percent confidence intervals; they do not incorporate the multiple-testing adjustment reported in Table S35. The dashed vertical line denotes equal effects. The horizontal scale runs from −1 to 1 baseline control-group standard deviations, without truncating the displayed intervals. A positive difference indicates a larger intervention effect in classrooms with female teachers, not the causal effect of teacher gender. This exploratory comparison across teaching contexts was not registered in the PAP and does not isolate a role-model mechanism. Source: survey data and the pooled interaction estimates reported in Table S35.

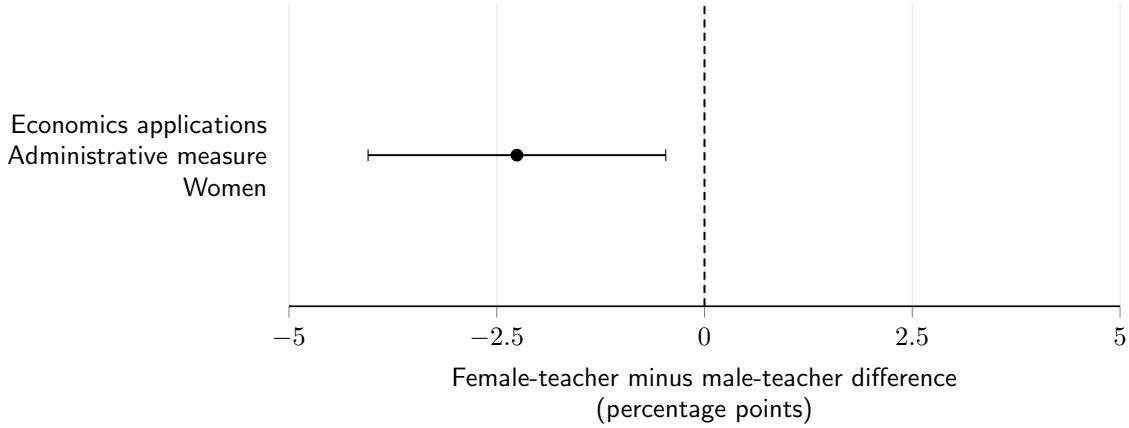



Figure S10: Economics-Application Effect Differences by Classroom Teacher Gender

*Notes:* The sample consists of women in the final year of four- and six-semester programs. The administrative application measure equals one for an application to Economics at Universidad Autónoma de Nuevo León or to the LEC economics program at Tecnológico de Monterrey, as recorded in the administrative registries. The binary outcome is multiplied by 100, so the plotted difference and its confidence interval are expressed in percentage points. The point is the female-teacher-minus-male-teacher difference in the treated-school assignment effect from a pooled model with assignment, teacher gender, and their interaction. Treated-school assignment combines directly treated and spillover classrooms and compares them with control schools. The specification includes the original randomization-block fixed effects and uses school-clustered standard errors. The horizontal bar is a pointwise, normal-approximation 95 percent confidence interval; it does not incorporate the multiple-testing adjustment reported in Table S35. The dashed vertical line denotes equal effects. The horizontal scale runs from −5 to 5 percentage points and displays the entire interval. This exploratory comparison across teaching contexts does not identify the effect of assigning a female teacher or isolate a role-model mechanism. It was not registered in the PAP. Sparse application events and limited cluster support warrant cautious interpretation. Source: administrative application records and the pooled interaction estimates reported in Table S35.

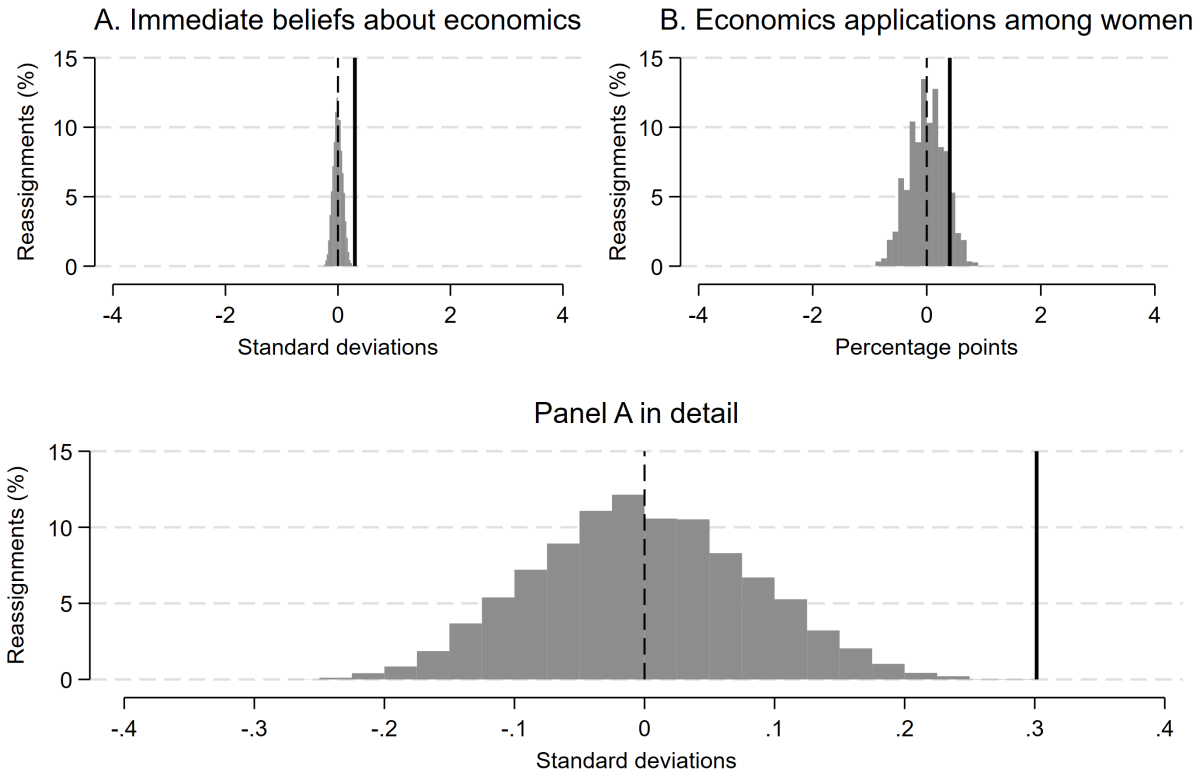



Figure S11: Randomization-Inference Null Distributions

*Notes:* Each histogram shows the percentage of 9,999 school reassignments in each bin under the sharp null of no treatment effect. Reassignments preserve the number of treated schools in each of the five randomization blocks. Panel A uses the immediate beliefs-about-economics index for final-semester students, in standard deviations relative to the baseline control group. Panel B uses applications to economics at UANL or Tec among final-semester women, in percentage points. Both panels use the same horizontal range. The lower plot enlarges Panel A around zero, using the same reassignments and bins. The dashed line marks zero and the solid line marks the observed effect. Including the observed assignment, the two-sided randomization-inference $p$-values are 0.0001 for immediate beliefs and 0.211 for applications.

Table S1: Immediate PAP Item Effects by Gender

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Wave: Immediate; Outcome family: Beliefs about economics; Outcome: Rejects money-and-finance stereotype* | | | | |
| Spillover vs. control | 0.155** | 0.023 | -0.131 | 0.130 |
| | (0.069) | (0.079) | (0.084) | |
| Treated-school assignment | 0.283*** | 0.176*** | -0.107 | 0.234 [0.265] |
| | (0.069) | (0.062) | (0.088) | |
| Direct treatment vs. control | 0.412*** | 0.349*** | -0.063 | 0.580 |
| | (0.108) | (0.072) | (0.112) | |
| Direct treatment vs. spillover | 0.263*** | 0.306*** | 0.043 | 0.707 |
| | (0.093) | (0.081) | (0.115) | |
| *Panel B. Wave: Immediate; Outcome family: Beliefs about economics; Outcome: Economics improves lives* | | | | |
| Spillover vs. control | -0.039 | -0.023 | 0.016 | 0.855 |
| | (0.073) | (0.085) | (0.084) | |
| Treated-school assignment | 0.024 | 0.160** | 0.136 | 0.206 [0.265] |
| | (0.106) | (0.076) | (0.105) | |
| Direct treatment vs. control | 0.086 | 0.359*** | 0.273* | 0.062 |
| | (0.166) | (0.092) | (0.140) | |
| Direct treatment vs. spillover | 0.119 | 0.409*** | 0.290** | 0.016 |
| | (0.104) | (0.097) | (0.118) | |
| *Panel C. Wave: Immediate; Outcome family: Beliefs about economics; Outcome: Understands what economists do* | | | | |
| Spillover vs. control | -0.063 | 0.023 | 0.086 | 0.236 |
| | (0.047) | (0.078) | (0.071) | |
| Treated-school assignment | 0.072 | 0.281*** | 0.209** | 0.019 [0.079] |
| | (0.092) | (0.084) | (0.083) | |
| Direct treatment vs. control | 0.198 | 0.563*** | 0.365*** | 0.002 |
| | (0.146) | (0.114) | (0.104) | |
| Direct treatment vs. spillover | 0.267*** | 0.582*** | 0.315*** | 0.002 |
| | (0.085) | (0.090) | (0.100) | |
| *Panel D. Wave: Immediate; Outcome family: Identity and self-concept; Outcome: Economics fits with identity* | | | | |
| Spillover vs. control | -0.027 | 0.053 | 0.080 | 0.155 |
| | (0.044) | (0.037) | (0.055) | |
| Treated-school assignment | 0.017 | 0.157*** | 0.140** | 0.011 [0.019] |
| | (0.042) | (0.030) | (0.051) | |
| Direct treatment vs. control | 0.061 | 0.274*** | 0.213** | 0.010 |
| | (0.072) | (0.047) | (0.077) | |
| Direct treatment vs. spillover | 0.093 | 0.232*** | 0.139* | 0.081 |
| | (0.068) | (0.049) | (0.079) | |

*Continued on next page*

Table S1 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel E. Wave: Immediate; Outcome family: Intentions and aspirations; Outcome: Likelihood of applying to economics* | | | | |
| Spillover vs. control | -0.047 | -0.001 | 0.047 | 0.448 |
| | (0.046) | (0.050) | (0.061) | |
| Treated-school assignment | -0.009 | 0.063 | 0.072* | 0.076 [0.182] |
| | (0.045) | (0.044) | (0.039) | |
| Direct treatment vs. control | 0.028 | 0.134* | 0.106 | 0.103 |
| | (0.053) | (0.070) | (0.063) | |
| Direct treatment vs. spillover | 0.073 | 0.130** | 0.057 | 0.473 |
| | (0.058) | (0.057) | (0.079) | |
| *Panel F. Wave: Immediate; Outcome family: Intentions and aspirations; Outcome: Family support for studying economics* | | | | |
| Spillover vs. control | 0.012 | -0.074 | -0.085 | 0.314 |
| | (0.076) | (0.065) | (0.083) | |
| Treated-school assignment | -0.038 | -0.067 | -0.029 | 0.721 [0.726] |
| | (0.099) | (0.066) | (0.081) | |
| Direct treatment vs. control | -0.083 | -0.060 | 0.023 | 0.820 |
| | (0.134) | (0.101) | (0.102) | |
| Direct treatment vs. spillover | -0.094 | 0.031 | 0.126 | 0.246 |
| | (0.081) | (0.095) | (0.107) | |
| *Panel G. Wave: Immediate; Outcome family: Gender stereotypes; Outcome: Rejects men-naturally-better stereotype* | | | | |
| Spillover vs. control | 0.119** | -0.003 | -0.121 | 0.128 |
| | (0.053) | (0.058) | (0.077) | |
| Treated-school assignment | 0.097* | 0.018 | -0.080 | 0.335 [0.364] |
| | (0.057) | (0.051) | (0.081) | |
| Direct treatment vs. control | 0.076 | 0.039 | -0.038 | 0.717 |
| | (0.085) | (0.056) | (0.103) | |
| Direct treatment vs. spillover | -0.030 | 0.052 | 0.082 | 0.472 |
| | (0.092) | (0.058) | (0.113) | |

*Notes:* The outcomes are immediate post-intervention items from the beliefs, identity, intentions, and gender-stereotype families. Each item is standardized using its baseline control-group mean and standard deviation, with the orientation indicated by its label. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. Regressions also control for the same item at baseline. The reported $p$-value tests the women-minus-men difference; bracketed $q$-values are Benjamini–Hochberg adjustments of wild-bootstrap tests for treated-school assignment within each original item family and wave. They are unavailable for the other contrasts, where no bracketed value is shown. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These gender-specific analyses are exploratory extensions of the study. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S2: Follow-Up PAP Item Effects by Gender: Beliefs about Economics

| Assignment contrast | Men | Women | Women - men | $p$-value [wild $q$-value] |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel A. Wave: Follow-up; Outcome family: Beliefs about economics; Outcome: Rejects money-and-finance stereotype* | | | | |
| Spillover vs. control | 0.167 | -0.029 | -0.196* | 0.056 |
| | (0.102) | (0.072) | (0.098) | |
| Treated-school assignment | 0.167** | 0.072 | -0.095 | 0.220 [0.268] |
| | (0.062) | (0.059) | (0.076) | |
| Direct treatment vs. control | 0.176* | 0.193** | 0.017 | 0.897 |
| | (0.096) | (0.091) | (0.131) | |
| Direct treatment vs. spillover | -0.032 | 0.207** | 0.238 | 0.113 |
| | (0.119) | (0.089) | (0.149) | |
| *Panel B. Wave: Follow-up; Outcome family: Beliefs about economics; Outcome: Economics improves lives* | | | | |
| Spillover vs. control | -0.104 | -0.004 | 0.100 | 0.233 |
| | (0.105) | (0.096) | (0.082) | |
| Treated-school assignment | 0.013 | 0.100 | 0.087 | 0.160 [0.268] |
| | (0.082) | (0.071) | (0.060) | |
| Direct treatment vs. control | 0.125 | 0.229** | 0.103 | 0.306 |
| | (0.103) | (0.083) | (0.099) | |
| Direct treatment vs. spillover | 0.310*** | 0.261** | -0.049 | 0.728 |
| | (0.106) | (0.102) | (0.140) | |

*Continued on next page*

Table S2 continued

| Assignment contrast | Men | Women | Women - men | $p$-value [wild $q$-value] |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel C. Wave: Follow-up; Outcome family: Beliefs about economics; Outcome: Understands what economists do* | | | | |
| Spillover vs. control | -0.007 | 0.056 | 0.064 | 0.549 |
| | (0.095) | (0.109) | (0.105) | |
| Treated-school assignment | 0.044 | 0.152 | 0.107 | 0.266 [0.268] |
| | (0.077) | (0.089) | (0.094) | |
| Direct treatment vs. control | 0.095 | 0.271** | 0.176 | 0.168 |
| | (0.101) | (0.112) | (0.124) | |
| Direct treatment vs. spillover | 0.206* | 0.228** | 0.022 | 0.891 |
| | (0.117) | (0.111) | (0.161) | |

*Notes:* The outcomes are follow-up items measuring beliefs about what economists do and the social value and financial orientation of economics. Each item is standardized using its baseline control-group mean and standard deviation, with the orientation indicated by its label. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. Regressions also control for the same item at baseline. The reported $p$-value tests the women-minus-men difference; bracketed $q$-values are Benjamini–Hochberg adjustments of wild-bootstrap tests for treated-school assignment within each original item family and wave. They are unavailable for the other contrasts, where no bracketed value is shown. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These gender-specific analyses are exploratory extensions of the study. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S3: Follow-Up PAP Item Effects by Gender: Identity and Self-Concept

| Assignment contrast | Men | Women | Women - men | $p$-value [wild $q$-value] |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel A. Wave: Follow-up; Outcome family: Identity and self-concept; Outcome: Self-assessed economics ability* | | | | |
| Spillover vs. control | 0.076 | -0.037 | -0.114 | 0.243 |
| | (0.068) | (0.080) | (0.095) | |
| Treated-school assignment | 0.046 | -0.023 | -0.068 | 0.406 [0.542] |
| | (0.067) | (0.070) | (0.081) | |
| Direct treatment vs. control | 0.019 | -0.008 | -0.027 | 0.799 |
| | (0.086) | (0.086) | (0.104) | |
| Direct treatment vs. spillover | 0.006 | 0.043 | 0.038 | 0.685 |
| | (0.078) | (0.073) | (0.092) | |
| *Panel B. Wave: Follow-up; Outcome family: Identity and self-concept; Outcome: Self-assessed mathematics ability* | | | | |
| Spillover vs. control | 0.026 | 0.031 | 0.005 | 0.939 |
| | (0.059) | (0.073) | (0.069) | |
| Treated-school assignment | -0.003 | 0.047 | 0.050 | 0.454 [0.542] |
| | (0.065) | (0.068) | (0.066) | |
| Direct treatment vs. control | -0.028 | 0.063 | 0.091 | 0.234 |
| | (0.084) | (0.077) | (0.075) | |
| Direct treatment vs. spillover | 0.005 | 0.065 | 0.060 | 0.430 |
| | (0.065) | (0.054) | (0.075) | |
| *Panel C. Wave: Follow-up; Outcome family: Identity and self-concept; Outcome: Feels comfortable considering economics* | | | | |
| Spillover vs. control | 0.058 | 0.115 | 0.058 | 0.566 |
| | (0.071) | (0.079) | (0.099) | |
| Treated-school assignment | 0.026 | 0.076 | 0.050 | 0.537 [0.570] |
| | (0.051) | (0.062) | (0.079) | |
| Direct treatment vs. control | -0.006 | 0.026 | 0.032 | 0.776 |
| | (0.073) | (0.072) | (0.111) | |
| Direct treatment vs. spillover | -0.079 | -0.103 | -0.024 | 0.843 |
| | (0.100) | (0.065) | (0.122) | |
| *Panel D. Wave: Follow-up; Outcome family: Identity and self-concept; Outcome: Ability relative to other-gender peers* | | | | |
| Spillover vs. control | 0.074 | 0.057 | -0.017 | 0.879 |
| | (0.076) | (0.084) | (0.107) | |
| Treated-school assignment | -0.035 | 0.083 | 0.118 | 0.206 [0.528] |
| | (0.069) | (0.081) | (0.091) | |
| Direct treatment vs. control | -0.134 | 0.103 | 0.238** | 0.040 |
| | (0.101) | (0.106) | (0.110) | |
| Direct treatment vs. spillover | -0.154 | 0.053 | 0.207 | 0.106 |
| | (0.100) | (0.084) | (0.127) | |

*Continued on next page*

Table S3 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel E. Wave: Follow-up; Outcome family: Identity and self-concept; Outcome: Ability relative to same-gender peers* | | | | |
| Spillover vs. control | 0.114 | 0.057 | -0.057 | 0.613 |
| | (0.072) | (0.078) | (0.111) | |
| Treated-school assignment | -0.012 | 0.065 | 0.077 | 0.430 [0.542] |
| | (0.076) | (0.071) | (0.096) | |
| Direct treatment vs. control | -0.129 | 0.063 | 0.192 | 0.123 |
| | (0.098) | (0.103) | (0.120) | |
| Direct treatment vs. spillover | -0.179* | 0.011 | 0.190 | 0.155 |
| | (0.094) | (0.095) | (0.132) | |
| *Panel F. Wave: Follow-up; Outcome family: Identity and self-concept; Outcome: Economics fits with identity* | | | | |
| Spillover vs. control | -0.101* | 0.025 | 0.127 | 0.135 |
| | (0.051) | (0.087) | (0.082) | |
| Treated-school assignment | -0.070 | 0.051 | 0.121* | 0.098 [0.528] |
| | (0.044) | (0.076) | (0.071) | |
| Direct treatment vs. control | -0.039 | 0.084 | 0.123 | 0.125 |
| | (0.052) | (0.076) | (0.078) | |
| Direct treatment vs. spillover | 0.076 | 0.080 | 0.005 | 0.961 |
| | (0.077) | (0.052) | (0.098) | |
| *Panel G. Wave: Follow-up; Outcome family: Identity and self-concept; Outcome: Can imagine an economist like me* | | | | |
| Spillover vs. control | -0.090 | -0.014 | 0.076 | 0.418 |
| | (0.057) | (0.068) | (0.093) | |
| Treated-school assignment | -0.086* | 0.009 | 0.095 | 0.145 [0.528] |
| | (0.044) | (0.057) | (0.064) | |
| Direct treatment vs. control | -0.080 | 0.037 | 0.117 | 0.108 |
| | (0.050) | (0.063) | (0.071) | |
| Direct treatment vs. spillover | 0.038 | 0.065 | 0.027 | 0.797 |
| | (0.082) | (0.057) | (0.106) | |

*Notes:* The outcomes are follow-up items measuring economics-related identity, self-concept, comfort, and perceived ability. Each item is standardized using its baseline control-group mean and standard deviation, with the orientation indicated by its label. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. Regressions also control for the same item at baseline. The reported $p$-value tests the women-minus-men difference; bracketed $q$-values are Benjamini–Hochberg adjustments of wild-bootstrap tests for treated-school assignment within each original item family and wave. They are unavailable for the other contrasts, where no bracketed value is shown. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These gender-specific analyses are exploratory extensions of the study. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S4: Follow-Up PAP Item Effects by Gender: Intentions and Stereotypes

| Assignment contrast | Men | Women | Women - men | $p$-value [wild $q$-value] |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel A. Wave: Follow-up; Outcome family: Intentions and aspirations; Outcome: Could afford studying economics* | | | | |
| Spillover vs. control | -0.009 | 0.013 | 0.022 | 0.880 |
| | (0.111) | (0.064) | (0.142) | |
| Treated-school assignment | 0.035 | 0.013 | -0.021 | 0.866 [0.968] |
| | (0.096) | (0.064) | (0.126) | |
| Direct treatment vs. control | 0.073 | 0.016 | -0.056 | 0.693 |
| | (0.111) | (0.078) | (0.141) | |
| Direct treatment vs. spillover | 0.096 | 0.006 | -0.090 | 0.547 |
| | (0.103) | (0.090) | (0.149) | |
| *Panel B. Wave: Follow-up; Outcome family: Intentions and aspirations; Outcome: Likelihood of applying to economics* | | | | |
| Spillover vs. control | -0.162** | 0.015 | 0.176* | 0.073 |
| | (0.059) | (0.072) | (0.095) | |
| Treated-school assignment | -0.099* | 0.015 | 0.115 | 0.104 [0.324] |
| | (0.052) | (0.056) | (0.068) | |
| Direct treatment vs. control | -0.044 | 0.021 | 0.065 | 0.454 |
| | (0.074) | (0.059) | (0.086) | |
| Direct treatment vs. spillover | 0.143* | 0.019 | -0.123 | 0.266 |
| | (0.085) | (0.064) | (0.110) | |
| *Panel C. Wave: Follow-up; Outcome family: Intentions and aspirations; Outcome: Family support for studying economics* | | | | |
| Spillover vs. control | 0.023 | 0.045 | 0.022 | 0.849 |
| | (0.109) | (0.069) | (0.115) | |
| Treated-school assignment | 0.026 | 0.021 | -0.005 | 0.967 [0.968] |
| | (0.093) | (0.062) | (0.118) | |
| Direct treatment vs. control | 0.026 | -0.007 | -0.034 | 0.811 |
| | (0.099) | (0.096) | (0.139) | |
| Direct treatment vs. spillover | -0.014 | -0.049 | -0.035 | 0.825 |
| | (0.123) | (0.108) | (0.157) | |
| *Panel D. Wave: Follow-up; Outcome family: Gender stereotypes; Outcome: Recognizes discrimination in economics* | | | | |
| Spillover vs. control | 0.031 | -0.136 | -0.166 | 0.177 |
| | (0.100) | (0.093) | (0.120) | |
| Treated-school assignment | 0.143 | -0.009 | -0.151 | 0.192 [0.424] |
| | (0.093) | (0.086) | (0.113) | |
| Direct treatment vs. control | 0.263** | 0.153 | -0.110 | 0.517 |
| | (0.107) | (0.123) | (0.168) | |
| Direct treatment vs. spillover | 0.284** | 0.330*** | 0.045 | 0.815 |
| | (0.133) | (0.099) | (0.192) | |

*Continued on next page*

Table S4 continued

| Assignment contrast | Men | Women | Women - men | $p$-value [wild $q$-value] |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel E. Wave: Follow-up; Outcome family: Gender stereotypes; Outcome: Rejects men-naturally-better stereotype* | | | | |
| Spillover vs. control | 0.190 | -0.010 | -0.199 | 0.106 |
| | (0.112) | (0.070) | (0.119) | |
| Treated-school assignment | 0.066 | 0.009 | -0.057 | 0.601 [0.624] |
| | (0.101) | (0.058) | (0.108) | |
| Direct treatment vs. control | -0.060 | 0.018 | 0.078 | 0.577 |
| | (0.125) | (0.078) | (0.138) | |
| Direct treatment vs. spillover | -0.259** | 0.010 | 0.269* | 0.083 |
| | (0.127) | (0.076) | (0.153) | |

*Notes:* The outcomes are follow-up items measuring intentions and aspirations (Panels A–C) and gender stereotypes (Panels D–E). Each item is standardized using its baseline control-group mean and standard deviation, with the orientation indicated by its label. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. Regressions also control for the same item at baseline. The reported $p$-value tests the women-minus-men difference; bracketed $q$-values are Benjamini–Hochberg adjustments of wild-bootstrap tests for treated-school assignment within each original item family and wave. They are unavailable for the other contrasts, where no bracketed value is shown. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These gender-specific analyses are exploratory extensions of the study. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S5: Gender Differences in Immediate Effects

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Outcome family: persistence; Outcome: Likelihood of applying to economics – immediate* | | | | |
| Spillover vs. control | -0.044 | 0.001 | 0.046 | 0.358 |
| | (0.059) | (0.056) | (0.049) | |
| Treated-school assignment | 0.046 | 0.062 | 0.016 | 0.725 [0.726] |
| | (0.059) | (0.048) | (0.044) | |
| Direct treatment vs. control | 0.129* | 0.139* | 0.010 | 0.898 |
| | (0.075) | (0.075) | (0.076) | |
| Direct treatment vs. spillover | 0.200** | 0.144** | -0.056 | 0.599 |
| | (0.087) | (0.069) | (0.106) | |
| *Panel B. Outcome family: persistence; Outcome: Family support for studying economics – immediate* | | | | |
| Spillover vs. control | 0.122* | -0.080 | -0.202** | 0.046 |
| | (0.061) | (0.081) | (0.097) | |
| Treated-school assignment | 0.105 | -0.037 | -0.142 | 0.105 [0.197] |
| | (0.074) | (0.069) | (0.085) | |
| Direct treatment vs. control | 0.095 | 0.011 | -0.083 | 0.443 |
| | (0.118) | (0.081) | (0.107) | |
| Direct treatment vs. spillover | -0.027 | 0.110 | 0.137 | 0.238 |
| | (0.117) | (0.076) | (0.115) | |
| *Panel C. Outcome family: persistence; Outcome: Rejects money-and-finance stereotype – immediate* | | | | |
| Spillover vs. control | 0.129 | -0.020 | -0.148 | 0.193 |
| | (0.091) | (0.083) | (0.111) | |
| Treated-school assignment | 0.225*** | 0.111 | -0.113 | 0.279 [0.395] |
| | (0.067) | (0.069) | (0.103) | |
| Direct treatment vs. control | 0.322** | 0.270*** | -0.052 | 0.743 |
| | (0.126) | (0.082) | (0.157) | |
| Direct treatment vs. spillover | 0.227* | 0.274*** | 0.047 | 0.788 |
| | (0.131) | (0.099) | (0.175) | |
| *Panel D. Outcome family: persistence; Outcome: Economics fits with identity – immediate* | | | | |
| Spillover vs. control | -0.064 | 0.050 | 0.115 | 0.109 |
| | (0.078) | (0.047) | (0.069) | |
| Treated-school assignment | 0.041 | 0.167*** | 0.126* | 0.065 [0.197] |
| | (0.066) | (0.048) | (0.066) | |
| Direct treatment vs. control | 0.143* | 0.310*** | 0.166* | 0.068 |
| | (0.082) | (0.064) | (0.088) | |
| Direct treatment vs. spillover | 0.213** | 0.271*** | 0.058 | 0.539 |
| | (0.085) | (0.053) | (0.094) | |
| *Panel E. Outcome family: persistence; Outcome: Rejects men-naturally-better stereotype – immediate* | | | | |
| Spillover vs. control | 0.134* | 0.006 | -0.128 | 0.171 |
| | (0.069) | (0.059) | (0.091) | |

*Continued on next page*

Table S5 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| Treated-school assignment | 0.117 | 0.010 | -0.107 | 0.313 [0.395] |
| | (0.079) | (0.045) | (0.104) | |
| Direct treatment vs. control | 0.099 | 0.012 | -0.086 | 0.566 |
| | (0.126) | (0.057) | (0.149) | |
| Direct treatment vs. spillover | 0.019 | 0.033 | 0.014 | 0.929 |
| | (0.131) | (0.073) | (0.161) | |
| *Panel F. Outcome family: persistence; Outcome: Economics improves lives – immediate* | | | | |
| Spillover vs. control | 0.058 | 0.102 | 0.045 | 0.677 |
| | (0.088) | (0.068) | (0.106) | |
| Treated-school assignment | 0.139* | 0.320*** | 0.181* | 0.069 [0.197] |
| | (0.080) | (0.074) | (0.096) | |
| Direct treatment vs. control | 0.223 | 0.564*** | 0.340** | 0.015 |
| | (0.153) | (0.097) | (0.130) | |
| Direct treatment vs. spillover | 0.209 | 0.490*** | 0.281* | 0.065 |
| | (0.156) | (0.088) | (0.150) | |
| *Panel G. Outcome family: persistence; Outcome: Understands what economists do – immediate* | | | | |
| Spillover vs. control | 0.036 | -0.021 | -0.058 | 0.619 |
| | (0.076) | (0.106) | (0.115) | |
| Treated-school assignment | 0.142 | 0.298*** | 0.156* | 0.097 [0.197] |
| | (0.102) | (0.097) | (0.090) | |
| Direct treatment vs. control | 0.248 | 0.668*** | 0.420*** | 0.002 |
| | (0.163) | (0.115) | (0.120) | |
| Direct treatment vs. spillover | 0.269** | 0.721*** | 0.451*** | 0.003 |
| | (0.131) | (0.100) | (0.148) | |

*Notes:* The dependent variables are the immediate levels of the seven items observed at both survey waves. These immediate effects use the same item-specific balanced samples as Table S6 and can differ from immediate estimates using all available respondents. The follow-up-minus-immediate change is tested directly in Table S7. All seven items retain the same baseline control-group standardization at both waves. For each item, the sample includes final-year baseline respondents with recorded gender and nonmissing baseline, immediate, and follow-up responses to that item. Thus, Tables S5, S6, and S7 use the same students for a given item, although samples differ across items. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. Regressions also include the baseline value of the item. The $p$-value tests the women-minus-men difference. Bracketed $q$-values adjust wild-bootstrap tests across the seven items separately for each wave or change specification and assignment contrast; bootstrap inference was computed only for treated-school assignment. No brackets are shown for contrasts without bootstrap results. These balanced-panel comparisons are exploratory and condition on response at both waves. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S6: Gender Differences in Follow-Up Effects

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Outcome family: persistence; Outcome: Likelihood of applying to economics – follow-up* | | | | |
| Spillover vs. control | -0.153*** | 0.026 | 0.179* | 0.055 |
| | (0.054) | (0.067) | (0.089) | |
| Treated-school assignment | -0.101** | 0.025 | 0.126* | 0.070 [0.194] |
| | (0.048) | (0.054) | (0.067) | |
| Direct treatment vs. control | -0.056 | 0.029 | 0.085 | 0.330 |
| | (0.072) | (0.057) | (0.085) | |
| Direct treatment vs. spillover | 0.121 | 0.018 | -0.103 | 0.340 |
| | (0.085) | (0.062) | (0.107) | |
| *Panel B. Outcome family: persistence; Outcome: Family support for studying economics – follow-up* | | | | |
| Spillover vs. control | -0.025 | -0.015 | 0.010 | 0.936 |
| | (0.115) | (0.067) | (0.127) | |
| Treated-school assignment | -0.006 | -0.028 | -0.021 | 0.867 [0.868] |
| | (0.098) | (0.061) | (0.126) | |
| Direct treatment vs. control | 0.009 | -0.041 | -0.050 | 0.745 |
| | (0.107) | (0.092) | (0.151) | |
| Direct treatment vs. spillover | 0.032 | -0.016 | -0.048 | 0.767 |
| | (0.125) | (0.104) | (0.163) | |
| *Panel C. Outcome family: persistence; Outcome: Rejects money-and-finance stereotype – follow-up* | | | | |
| Spillover vs. control | 0.193** | -0.005 | -0.198** | 0.040 |
| | (0.094) | (0.080) | (0.091) | |
| Treated-school assignment | 0.169** | 0.113* | -0.056 | 0.480 [0.574] |
| | (0.062) | (0.063) | (0.079) | |
| Direct treatment vs. control | 0.158 | 0.247** | 0.089 | 0.485 |
| | (0.095) | (0.090) | (0.126) | |
| Direct treatment vs. spillover | -0.074 | 0.231** | 0.305** | 0.042 |
| | (0.115) | (0.093) | (0.148) | |
| *Panel D. Outcome family: persistence; Outcome: Economics fits with identity – follow-up* | | | | |
| Spillover vs. control | -0.100* | 0.039 | 0.140* | 0.074 |
| | (0.056) | (0.080) | (0.075) | |
| Treated-school assignment | -0.067 | 0.066 | 0.133* | 0.059 [0.194] |
| | (0.044) | (0.073) | (0.067) | |
| Direct treatment vs. control | -0.035 | 0.099 | 0.134* | 0.096 |
| | (0.050) | (0.076) | (0.078) | |
| Direct treatment vs. spillover | 0.078 | 0.083 | 0.005 | 0.959 |
| | (0.075) | (0.050) | (0.094) | |
| *Panel E. Outcome family: persistence; Outcome: Rejects men-naturally-better stereotype – follow-up* | | | | |
| Spillover vs. control | 0.156 | -0.046 | -0.202 | 0.130 |
| | (0.121) | (0.066) | (0.130) | |

*Continued on next page*

Table S6 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| Treated-school assignment | 0.069 | -0.029 | -0.098 | 0.402 [0.574] |
| | (0.110) | (0.056) | (0.115) | |
| Direct treatment vs. control | -0.023 | -0.018 | 0.005 | 0.974 |
| | (0.134) | (0.081) | (0.142) | |
| Direct treatment vs. spillover | -0.172 | 0.009 | 0.181 | 0.259 |
| | (0.133) | (0.079) | (0.159) | |
| *Panel F. Outcome family: persistence; Outcome: Economics improves lives – follow-up* | | | | |
| Spillover vs. control | -0.118 | -0.012 | 0.107 | 0.281 |
| | (0.117) | (0.099) | (0.097) | |
| Treated-school assignment | 0.034 | 0.099 | 0.066 | 0.292 [0.511] |
| | (0.079) | (0.073) | (0.061) | |
| Direct treatment vs. control | 0.168 | 0.231** | 0.063 | 0.526 |
| | (0.100) | (0.089) | (0.098) | |
| Direct treatment vs. spillover | 0.375*** | 0.279*** | -0.096 | 0.506 |
| | (0.110) | (0.101) | (0.143) | |
| *Panel G. Outcome family: persistence; Outcome: Understands what economists do – follow-up* | | | | |
| Spillover vs. control | -0.041 | 0.101 | 0.142 | 0.170 |
| | (0.091) | (0.104) | (0.100) | |
| Treated-school assignment | 0.008 | 0.166* | 0.158* | 0.078 [0.194] |
| | (0.072) | (0.086) | (0.086) | |
| Direct treatment vs. control | 0.050 | 0.244** | 0.194 | 0.135 |
| | (0.103) | (0.109) | (0.126) | |
| Direct treatment vs. spillover | 0.202* | 0.152 | -0.051 | 0.750 |
| | (0.120) | (0.107) | (0.159) | |

*Notes:* The dependent variables are the follow-up levels of the seven items observed at both survey waves. These are balanced-panel follow-up effects, not follow-up-minus-immediate changes. The latter are tested directly in Table S7; different significance levels across waves do not establish a change. All seven items retain the same baseline control-group standardization at both waves. For each item, the sample includes final-year baseline respondents with recorded gender and nonmissing baseline, immediate, and follow-up responses to that item. Thus, Tables S5, S6, and S7 use the same students for a given item, although samples differ across items. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. Regressions also include the baseline value of the item. The $p$-value tests the women-minus-men difference. Bracketed $q$-values adjust wild-bootstrap tests across the seven items separately for each wave or change specification and assignment contrast; bootstrap inference was computed only for treated-school assignment. No brackets are shown for contrasts without bootstrap results. These balanced-panel comparisons are exploratory and condition on response at both waves. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S7: Gender Differences in Persistence: Follow-Up Minus Immediate

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Outcome family: persistence; Outcome: Likelihood of applying to economics – follow-up minus immediate* | | | | |
| Spillover vs. control | -0.109 | 0.025 | 0.133 | 0.126 |
| | (0.074) | (0.069) | (0.084) | |
| Treated-school assignment | -0.147** | -0.036 | 0.110* | 0.071 [0.416] |
| | (0.059) | (0.058) | (0.059) | |
| Direct treatment vs. control | -0.185** | -0.110 | 0.075 | 0.399 |
| | (0.075) | (0.073) | (0.087) | |
| Direct treatment vs. spillover | -0.079 | -0.126* | -0.047 | 0.706 |
| | (0.095) | (0.073) | (0.124) | |
| *Panel B. Outcome family: persistence; Outcome: Family support for studying economics – follow-up minus immediate* | | | | |
| Spillover vs. control | -0.147 | 0.065 | 0.212 | 0.239 |
| | (0.112) | (0.110) | (0.176) | |
| Treated-school assignment | -0.112 | 0.009 | 0.121 | 0.473 [0.976] |
| | (0.118) | (0.095) | (0.166) | |
| Direct treatment vs. control | -0.086 | -0.052 | 0.034 | 0.860 |
| | (0.152) | (0.105) | (0.190) | |
| Direct treatment vs. spillover | 0.059 | -0.126 | -0.185 | 0.363 |
| | (0.178) | (0.128) | (0.202) | |
| *Panel C. Outcome family: persistence; Outcome: Rejects money-and-finance stereotype – follow-up minus immediate* | | | | |
| Spillover vs. control | 0.064 | 0.015 | -0.049 | 0.634 |
| | (0.079) | (0.063) | (0.102) | |
| Treated-school assignment | -0.056 | 0.002 | 0.057 | 0.622 [0.976] |
| | (0.078) | (0.067) | (0.114) | |
| Direct treatment vs. control | -0.164 | -0.023 | 0.141 | 0.436 |
| | (0.128) | (0.101) | (0.179) | |
| Direct treatment vs. spillover | -0.301** | -0.043 | 0.258 | 0.202 |
| | (0.133) | (0.118) | (0.200) | |
| *Panel D. Outcome family: persistence; Outcome: Economics fits with identity – follow-up minus immediate* | | | | |
| Spillover vs. control | -0.036 | -0.011 | 0.025 | 0.814 |
| | (0.072) | (0.076) | (0.105) | |
| Treated-school assignment | -0.108 | -0.101 | 0.007 | 0.938 [0.976] |
| | (0.066) | (0.063) | (0.084) | |
| Direct treatment vs. control | -0.179* | -0.211** | -0.032 | 0.749 |
| | (0.087) | (0.077) | (0.100) | |
| Direct treatment vs. spillover | -0.135 | -0.188*** | -0.053 | 0.647 |
| | (0.091) | (0.063) | (0.116) | |

*Continued on next page*

Table S7 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel E. Outcome family: persistence; Outcome: Rejects men-naturally-better stereotype – follow-up minus immediate* | | | | |
| Spillover vs. control | 0.022 | -0.053 | -0.074 | 0.569 |
| | (0.113) | (0.094) | (0.129) | |
| Treated-school assignment | -0.048 | -0.039 | 0.009 | 0.931 [0.976] |
| | (0.111) | (0.065) | (0.104) | |
| Direct treatment vs. control | -0.122 | -0.031 | 0.091 | 0.544 |
| | (0.153) | (0.105) | (0.148) | |
| Direct treatment vs. spillover | -0.191 | -0.024 | 0.167 | 0.362 |
| | (0.148) | (0.100) | (0.182) | |
| *Panel F. Outcome family: persistence; Outcome: Economics improves lives – follow-up minus immediate* | | | | |
| Spillover vs. control | -0.176 | -0.114 | 0.062 | 0.458 |
| | (0.110) | (0.115) | (0.082) | |
| Treated-school assignment | -0.105 | -0.221** | -0.115 | 0.194 [0.723] |
| | (0.103) | (0.094) | (0.087) | |
| Direct treatment vs. control | -0.056 | -0.333*** | -0.278* | 0.070 |
| | (0.159) | (0.094) | (0.147) | |
| Direct treatment vs. spillover | 0.166 | -0.211* | -0.377** | 0.037 |
| | (0.157) | (0.108) | (0.178) | |

*Continued on next page*

Table S7 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel G. Outcome family: persistence; Outcome: Understands what economists do – follow-up minus immediate* | | | | |
| Spillover vs. control | -0.077 | 0.122 | 0.199* | 0.093 |
| | (0.113) | (0.098) | (0.114) | |
| Treated-school assignment | -0.134 | -0.132 | 0.003 | 0.976 [0.976] |
| | (0.096) | (0.079) | (0.090) | |
| Direct treatment vs. control | -0.197 | -0.423*** | -0.226* | 0.087 |
| | (0.133) | (0.084) | (0.127) | |
| Direct treatment vs. spillover | -0.067 | -0.569*** | -0.502*** | 0.003 |
| | (0.148) | (0.091) | (0.162) | |

*Notes:* The dependent variable is the within-student change in each item, defined as follow-up minus immediate response. Men and Women therefore measure changes in treatment effects, and Women - men tests whether that change differs by gender. A positive difference indicates that the treatment effect evolved more positively for women than for men; it does not imply that either follow-up effect is itself positive. All seven items retain the same baseline control-group standardization at both waves. For each item, the sample includes final-year baseline respondents with recorded gender and nonmissing baseline, immediate, and follow-up responses to that item. Thus, Tables S5, S6, and S7 use the same students for a given item, although samples differ across items. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. Regressions also include the baseline value of the item. The $p$-value tests the women-minus-men difference. Bracketed $q$-values adjust wild-bootstrap tests across the seven items separately for each wave or change specification and assignment contrast; bootstrap inference was computed only for treated-school assignment. No brackets are shown for contrasts without bootstrap results. These balanced-panel comparisons are exploratory and condition on response at both waves. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S8: Economics Enrollment and UANL Acceptance: School-Assignment Contrasts

| Outcome | Spillover vs. control (1) | Treated-school assignment (2) | Direct treatment vs. control (3) | Covariate-adjusted school assignment (4) | Control mean (5) |
|---|---|---|---|---|---|
| *Panel A. Women in the final semester* | | | | | |
| Enrollment in an economics degree | -0.292* | -0.014 | 0.312 | -0.014 | 0.196 |
| | (0.167) [0.093] | (0.169) [0.934] | (0.279) [0.273] | (0.171) [0.935] | |
| Enrollment in economics: broad | -0.449* | 0.093 | 0.729* | 0.077 | 0.392 |
| | (0.239) [0.070] | (0.244) [0.705] | (0.426) [0.098] | (0.223) [0.733] | |
| UANL economics acceptance | 0.416 | 0.348 | 0.269 | 0.360 | 0.098 |
| | (0.306) [0.185] | (0.230) [0.141] | (0.253) [0.298] | (0.236) [0.139] | |
| Enrollment in Tec ECSG gateway | -0.157 | 0.107 | 0.417 | 0.091 | 0.196 |
| | (0.142) [0.277] | (0.185) [0.567] | (0.326) [0.212] | (0.170) [0.597] | |
| *Panel B. All final-semester students* | | | | | |
| Enrollment in an economics degree | -0.270 | -0.124 | 0.032 | -0.056 | 0.472 |
| | (0.223) [0.237] | (0.216) [0.570] | (0.331) [0.924] | (0.212) [0.793] | |
| Enrollment in economics: broad | -0.347 | -0.025 | 0.321 | 0.054 | 0.577 |
| | (0.234) [0.149] | (0.213) [0.909] | (0.343) [0.357] | (0.217) [0.803] | |
| UANL economics acceptance | 0.170 | 0.346* | 0.534 | 0.379** | 0.210 |
| | (0.247) [0.497] | (0.183) [0.070] | (0.369) [0.159] | (0.173) [0.037] | |
| Enrollment in Tec ECSG gateway | -0.275* | -0.107 | 0.072 | -0.083 | 0.210 |
| | (0.160) [0.096] | (0.150) [0.482] | (0.232) [0.757] | (0.140) [0.557] | |

*Notes:* Economics-degree enrollment combines Tecnológico de Monterrey (Tec) Licenciatura en Economía (LEC) enrollment records with August-survey reports naming an economics degree. Broad enrollment additionally includes enrollment in Tec's Ciencias Sociales y Gobierno (ECSG) gateway; it is not enrollment in an economics degree alone. The Universidad Autónoma de Nuevo León (UANL) outcome records acceptance to Economics, not matriculation, which is not observed in that registry. The last row records Tec ECSG enrollment. Outcomes are indicators multiplied by 100; coefficients and standard errors are in percentage points and comparison-group means are unadjusted percentages. Missing evidence is coded as no detected event, so incomplete August-survey coverage limits the mixed-source enrollment measures. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Spillover versus control compares spillover classrooms in treated schools with control schools; direct treatment versus control compares direct-treatment classrooms with control schools. All specifications include the original randomization-block fixed effects and use school-clustered standard errors. The covariate-adjusted school-assignment specification additionally includes baseline age, GPA, economics application probability, interest in economics, and self-assessed mathematics ability, with mean imputation and missing-value indicators. The other specifications contain no additional covariates. Panel A includes women in the final semester and Panel B all final-semester students, restricted in both cases to the 2026 baseline sample. Effect cells report the coefficient, its standard error in parentheses, and the two-sided $p$-value in brackets; these brackets do not contain $q$-values. *, **, and *** denote two-sided conventional $p$-values below 0.10, 0.05, and 0.01, respectively. These administrative-stage extensions are exploratory, and sparse events warrant caution. Sources are the linked institutional records and August student survey. Significance is assessed using two-sided, unrounded $p$-values below 0.10.

Table S9: Economics Enrollment and UANL Acceptance: Within-Treated-School Contrast

| Outcome | Direct treatment vs. spillover (1) | Spillover mean (2) |
|---|---|---|
| *Panel A. Women in the final semester* | | |
| Enrollment in an economics degree | 0.559* | 0.000 |
| | (0.331) [0.095] | |
| Enrollment in economics: broad | 1.009* | 0.579 |
| | (0.533) [0.062] | |
| UANL economics acceptance | -0.041 | 0.386 |
| | (0.297) [0.891] | |
| Enrollment in Tec ECSG gateway | 0.451 | 0.579 |
| | (0.463) [0.333] | |
| *Panel B. All final-semester students* | | |
| Enrollment in an economics degree | 0.278 | 0.426 |
| | (0.369) [0.454] | |
| Enrollment in economics: broad | 0.593 | 0.852 |
| | (0.397) [0.139] | |
| UANL economics acceptance | 0.404 | 0.319 |
| | (0.372) [0.281] | |
| Enrollment in Tec ECSG gateway | 0.315 | 0.426 |
| | (0.241) [0.194] | |

*Notes:* Economics-degree enrollment combines Tecnológico de Monterrey (Tec) Licenciatura en Economía (LEC) enrollment records with August-survey reports naming an economics degree. Broad enrollment additionally includes enrollment in Tec's Ciencias Sociales y Gobierno (ECSG) gateway; it is not enrollment in an economics degree alone. The Universidad Autónoma de Nuevo León (UANL) outcome records acceptance to Economics, not matriculation, which is not observed in that registry. The last row records Tec ECSG enrollment. Outcomes are indicators multiplied by 100; coefficients and standard errors are in percentage points and comparison-group means are unadjusted percentages. Missing evidence is coded as no detected event, so incomplete August-survey coverage limits the mixed-source enrollment measures. Direct treatment versus spillover compares direct-treatment with spillover classrooms within treated schools, includes school fixed effects, and uses classroom-clustered standard errors, without additional covariates. Panel A includes women in the final semester and Panel B all final-semester students, restricted in both cases to the 2026 baseline sample. Both panels are further restricted to treated schools; the reported mean is for their spillover classrooms. Effect cells report the coefficient, its standard error in parentheses, and the two-sided $p$-value in brackets; these brackets do not contain $q$-values. *, **, and *** denote two-sided conventional $p$-values below 0.10, 0.05, and 0.01, respectively. These administrative-stage extensions are exploratory, and sparse events warrant caution. Sources are the linked institutional records and August student survey. Significance is assessed using two-sided, unrounded $p$-values below 0.10.

Table S10: Application, Admission, and Enrollment Outcomes by Gender

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Outcome family: Applications; Outcome: Combined application measure* | | | | |
| Spillover vs. control | 0.252 | 0.028 | -0.224 | 0.760 [0.764] |
| | (0.677) | (0.475) | (0.725) | |
| Treated-school assignment | 0.912 | 0.144 | -0.769 | 0.393 [0.693] |
| | (0.717) | (0.435) | (0.886) | |
| Direct treatment vs. control | 1.549 | 0.291 | -1.259 | 0.399 [0.806] |
| | (1.243) | (0.486) | (1.469) | |
| Direct treatment vs. spillover | 1.262 | 0.206 | -1.056 | 0.412 [0.460] |
| | (1.191) | (0.562) | (1.280) | |
| *Panel B. Outcome family: Applications; Outcome: Self-reported economics outside both registries* | | | | |
| Spillover vs. control | 0.180 | -0.055 | -0.234 | 0.335 |
| | (0.214) | (0.057) | (0.239) | |
| Treated-school assignment | 0.166 | 0.039 | -0.127 | 0.531 |
| | (0.145) | (0.083) | (0.200) | |
| Direct treatment vs. control | 0.156 | 0.145 | -0.010 | 0.975 |
| | (0.238) | (0.187) | (0.327) | |
| Direct treatment vs. spillover | -0.017 | 0.153 | 0.170 | 0.660 |
| | (0.323) | (0.219) | (0.385) | |
| *Panel C. Outcome family: Applications; Outcome: Tec ECSG gateway application* | | | | |
| Spillover vs. control | -0.825 | -0.077 | 0.749 | 0.147 |
| | (0.506) | (0.256) | (0.502) | |
| Treated-school assignment | -0.888 | 0.191 | 1.078* | 0.097 |
| | (0.562) | (0.277) | (0.627) | |
| Direct treatment vs. control | -0.941 | 0.493 | 1.433* | 0.088 |
| | (0.658) | (0.386) | (0.811) | |
| Direct treatment vs. spillover | -0.073 | 0.479 | 0.553 | 0.546 |
| | (0.461) | (0.555) | (0.912) | |
| *Panel D. Outcome family: Applications; Outcome: Enrollment in an economics degree* | | | | |
| Spillover vs. control | -0.057 | -0.429** | -0.372 | 0.546 |
| | (0.524) | (0.192) | (0.608) | |
| Treated-school assignment | -0.107 | -0.147 | -0.040 | 0.938 |
| | (0.436) | (0.212) | (0.507) | |
| Direct treatment vs. control | -0.148 | 0.173 | 0.321 | 0.658 |
| | (0.629) | (0.317) | (0.717) | |
| Direct treatment vs. spillover | -0.077 | 0.538 | 0.615 | 0.417 |
| | (0.684) | (0.354) | (0.754) | |

*Continued on next page*

Table S10 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel E. Outcome family: Applications; Outcome: Enrolled at Tec in ECSG gateway* | | | | |
| Spillover vs. control | -0.496 | -0.053 | 0.443 | 0.228 |
| | (0.321) | (0.138) | (0.359) | |
| Treated-school assignment | -0.535 | 0.227 | 0.761 | 0.125 |
| | (0.321) | (0.260) | (0.480) | |
| Direct treatment vs. control | -0.564* | 0.544 | 1.108* | 0.090 |
| | (0.322) | (0.432) | (0.631) | |
| Direct treatment vs. spillover | -0.036 | 0.564 | 0.600 | 0.392 |
| | (0.345) | (0.472) | (0.697) | |
| *Panel F. Outcome family: Applications; Outcome: Admitted to Tec, any program* | | | | |
| Spillover vs. control | -0.879 | 1.166 | 2.045 | 0.242 |
| | (1.103) | (1.023) | (1.709) | |
| Treated-school assignment | 0.550 | 1.127 | 0.577 | 0.740 |
| | (1.345) | (0.935) | (1.723) | |
| Direct treatment vs. control | 1.920 | 1.117 | -0.803 | 0.716 |
| | (1.953) | (0.955) | (2.186) | |
| Direct treatment vs. spillover | 2.487* | 0.798 | -1.689 | 0.386 |
| | (1.364) | (1.159) | (1.938) | |

*Notes:* All outcomes are binary indicators multiplied by 100; coefficients and standard errors are in percentage points. Panel A combines economics applications detected in institutional records or the August survey; Panel B captures self-reported economics applications outside the Universidad Autónoma de Nuevo León (UANL) and Tecnológico de Monterrey (Tec) registries. Panel C records application to Tec's Ciencias Sociales y Gobierno (ECSG) gateway. Economics-degree enrollment in Panel D combines Tec Licenciatura en Economía (LEC) enrollment with August-survey economics-degree reports; Panel E records ECSG enrollment, and Panel F admission to any Tec program. These stages and definitions are not interchangeable. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No model conditions on August-survey response or includes additional covariates beyond the assignment and gender terms described above. The $p$-value tests the gender difference. Bracketed wild-bootstrap $q$-values appear only for Panel A and retain the Benjamini–Hochberg adjustment across the combined, administrative, and gateway application definitions in the original analysis, separately by contrast. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. Missing event evidence is coded as no detected event, not verified nonparticipation; mixed-source outcomes are sensitive to incomplete August-survey coverage. Panel D is detected mixed-source enrollment, not a complete administrative enrollment measure. Sparse events and reporting require cautious interpretation. These stage-by-gender extensions are exploratory. Sources are linked institutional records and the August student survey. Standard errors 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 S11: Predetermined Covariates of Economics Applications

| Sample | Outcome (1) | Covariate (2) | Association (3) | $p$-value (4) | Sample size (5) |
|---|---|---|---|---|---|
| *Panel A. Sample: Women in the final semester; Outcome: Combined application measure* | | | | | |
| Women in the final semester | Combined application measure | Age | 0.155 | 0.483 | 2,009 |
| | | | (0.219) | | |
| Women in the final semester | Combined application measure | Baseline GPA | 0.047 | 0.162 | 2,009 |
| | | | (0.033) | | |
| Women in the final semester | Combined application measure | Likelihood of applying to economics | 0.313*** | 0.007 | 2,009 |
| | | | (0.106) | | |
| Women in the final semester | Combined application measure | Interest in economics | 0.065** | 0.037 | 2,009 |
| | | | (0.030) | | |
| Women in the final semester | Combined application measure | Self-assessed mathematics ability | -0.014 | 0.668 | 2,009 |
| | | | (0.032) | | |
| *Panel B. Sample: Women in the final semester; Outcome: Administrative applications* | | | | | |
| Women in the final semester | Administrative applications | Age | 0.291 | 0.181 | 2,009 |
| | | | (0.212) | | |
| Women in the final semester | Administrative applications | Baseline GPA | 0.029 | 0.257 | 2,009 |
| | | | (0.025) | | |
| Women in the final semester | Administrative applications | Likelihood of applying to economics | 0.274** | 0.010 | 2,009 |
| | | | (0.099) | | |
| Women in the final semester | Administrative applications | Interest in economics | 0.074** | 0.019 | 2,009 |
| | | | (0.030) | | |
| Women in the final semester | Administrative applications | Self-assessed mathematics ability | -0.004 | 0.883 | 2,009 |
| | | | (0.027) | | |
| *Panel C. Sample: All final-semester students; Outcome: Combined application measure* | | | | | |
| All final-semester students | Combined application measure | Age | 0.092 | 0.602 | 3,763 |
| | | | (0.175) | | |
| All final-semester students | Combined application measure | Baseline GPA | 0.030 | 0.252 | 3,763 |
| | | | (0.026) | | |
| All final-semester students | Combined application measure | Likelihood of applying to economics | 0.517*** | $<$ 0.001 | 3,763 |
| | | | (0.134) | | |
| All final-semester students | Combined application measure | Interest in economics | 0.130*** | $<$ 0.001 | 3,763 |
| | | | (0.032) | | |

*Continued on next page*

Table S11 continued

| Sample | Outcome (1) | Covariate (2) | Association (3) | $p$-value (4) | Sample size (5) |
|---|---|---|---|---|---|
| All final-semester students | Combined application measure | Self-assessed mathematics ability | -0.029 (0.042) | 0.493 | 3,763 |
| *Panel D. Sample: All final-semester students; Outcome: Administrative applications* | | | | | |
| All final-semester students | Administrative applications | Age | -0.038 (0.171) | 0.825 | 3,763 |
| All final-semester students | Administrative applications | Baseline GPA | 0.020 (0.022) | 0.369 | 3,763 |
| All final-semester students | Administrative applications | Likelihood of applying to economics | 0.411*** (0.113) | 0.001 | 3,763 |
| All final-semester students | Administrative applications | Interest in economics | 0.117*** (0.025) | < 0.001 | 3,763 |
| All final-semester students | Administrative applications | Self-assessed mathematics ability | -0.033 (0.039) | 0.401 | 3,763 |

*Notes:* Panels distinguish women and all final-semester students and the combined and administrative application measures. Each coefficient is the partial association between a one-unit increase in the named baseline covariate and the application outcome, estimated jointly with treated-school assignment, the other listed baseline covariates, and the original randomization-block fixed effects. The model for the combined measure records detected events in the full eligible sample and does not condition on August-survey response; incomplete survey coverage remains a limitation of that outcome. Application outcomes are binary indicators multiplied by 100; effects, standard errors, and control means are therefore reported in percentage points. Standard errors are clustered by school. Because the covariates are not randomized, these coefficients summarize conditional predictiveness and should not be interpreted as causal effects of age, GPA, application propensity, interest, or mathematics confidence. Standard errors 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 S12: Reported Attendance and Reception of the Visit by Gender

| Study arm or survey item | Men Mean (1) | Women Mean (2) | Men Valid response (%) (3) | Women Valid response (%) (4) |
|---|---|---|---|---|
| *Panel A. Reported attendance at the female-economist session (%)* | | | | |
| Control | 44.30 | 43.64 | 72.6 | 77.1 |
| Spillover | 66.14 | 59.32 | 78.4 | 79.7 |
| Direct treatment | 88.50 | 91.91 | 93.5 | 94.5 |
| *Panel B. Reception in directly assigned classrooms (0–10)* | | | | |
| Can recall the speaker's name or affiliation | 7.20 | 7.58 | 83.4 | 86.4 |
| Enjoyed the session | 8.17 | 8.56 | 83.4 | 86.6 |
| Visit increased interest in economics | 5.89 | 5.63 | 83.4 | 86.6 |
| Visit increased confidence in quantitative ability | 6.51 | 6.48 | 82.7 | 86.4 |
| After the visit, can imagine being an economist | 5.15 | 4.66 | 83.2 | 86.6 |
| Perception of economists' work changed | 7.24 | 7.24 | 82.5 | 86.4 |

*Notes:* Panel A reports affirmative answers to the question asking whether the student attended the session in which a female economist visited the class, as percentages of valid yes/no responses. Control denotes classrooms in control schools; spillover denotes classrooms assigned the active placebo in treated schools; direct treatment denotes classrooms assigned the role-model session. These answers are self-reports and are not independently verified attendance. Affirmative answers outside the direct-treatment arm may reflect differences in how students understood the question. Panel B is restricted to students in directly assigned classrooms with a valid response to each item. Entries are unweighted means on scales from 0 (strongly disagree) to 10 (strongly agree). The statements record enjoyment, recall, and perceived changes attributed to the visit; reporting a change in perceptions does not establish its direction. The sample comprises final-year baseline respondents from both four- and six-semester programs with recorded gender. Valid response (%) is the number with a valid answer divided by all eligible students of the same gender in the corresponding arm, multiplied by 100. For every Panel B row, the denominator is the eligible direct-treatment group of that gender. Missing, not-applicable, and out-of-range responses are excluded from the means, without being recoded to zero. Panel B does not additionally condition on reported attendance. The table contains descriptive means and response coverage, with no regression adjustment or significance tests. Gender differences in these selected responses do not identify differential treatment effects or causal mediation. The data come from the study's baseline and subsequent student surveys.

Table S13: Recall, Peer Discussion, and Family Diffusion

| Assignment contrast | Effect | $p$-value [$q$-value] | Control mean | Spillover mean | Direct-treatment mean |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Panel A. Reports a female-economist activity* | | | | | |
| Spillover vs. control | -1.101 | 0.683 [0.777] | 90.252 | 88.008 | 85.577 |
| | (2.670) | | | | |
| Direct treatment vs. control | -2.825 | 0.363 [0.528] | 90.252 | 88.008 | 85.577 |
| | (3.055) | | | | |
| Direct treatment vs. spillover | -0.027 | 0.995 [0.995] | 90.252 | 88.008 | 85.577 |
| | (4.163) | | | | |
| *Panel B. Discussed economics with family, friends, or peers* | | | | | |
| Spillover vs. control | 0.050 | 0.432 [0.693] | 3.564 | 3.870 | 3.986 |
| | (0.063) | | | | |
| Direct treatment vs. control | 0.059 | 0.396 [0.528] | 3.564 | 3.870 | 3.986 |
| | (0.069) | | | | |
| Direct treatment vs. spillover | 0.067 | 0.366 [0.488] | 3.564 | 3.870 | 3.986 |
| | (0.074) | | | | |
| *Panel C. Can name a female economist* | | | | | |
| Spillover vs. control | 4.099 | 0.197 [0.693] | 17.188 | 20.200 | 26.306 |
| | (3.100) | | | | |
| Direct treatment vs. control | 10.371*** | 0.010 [0.026] | 17.188 | 20.200 | 26.306 |
| | (3.729) | | | | |
| Direct treatment vs. spillover | 5.159 | 0.100 [0.200] | 17.188 | 20.200 | 26.306 |
| | (3.100) | | | | |
| *Panel D. Someone discussed studying economics with student* | | | | | |
| Spillover vs. control | -0.046 | 0.537 [0.715] | 3.547 | 3.488 | 3.694 |
| | (0.073) | | | | |
| Direct treatment vs. control | -0.001 | 0.982 [0.982] | 3.547 | 3.488 | 3.694 |
| | (0.064) | | | | |
| Direct treatment vs. spillover | 0.111* | 0.070 [0.187] | 3.547 | 3.488 | 3.694 |
| | (0.060) | | | | |
| *Panel E. Any or partial recall of visiting economist* | | | | | |
| Spillover vs. control | 17.218 | 0.231 [0.693] | 43.702 | 56.286 | 90.385 |
| | (14.042) | | | | |
| Direct treatment vs. control | 53.754*** | $<$ 0.001 [$<$ 0.001] | 43.702 | 56.286 | 90.385 |
| | (11.989) | | | | |
| Direct treatment vs. spillover | 34.408*** | $<$ 0.001 [$<$ 0.001] | 43.702 | 56.286 | 90.385 |
| | (6.663) | | | | |

*Continued on next page*

Table S13 continued

| Assignment contrast | Effect | $p$-value [$q$-value] | Control mean | Spillover mean | Direct-treatment mean |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Panel F. Exact recall of visiting economist* | | | | | |
| Spillover vs. control | 15.654 | 0.262 [0.693] | 40.854 | 51.838 | 88.846 |
| | (13.675) | | | | |
| Direct treatment vs. control | 55.137*** | $<$ 0.001 [$<$ 0.001] | 40.854 | 51.838 | 88.846 |
| | (11.763) | | | | |
| Direct treatment vs. spillover | 37.213*** | $<$ 0.001 [$<$ 0.001] | 40.854 | 51.838 | 88.846 |
| | (7.082) | | | | |
| *Panel G. Can describe at least one topic from activity* | | | | | |
| Spillover vs. control | -0.017 | 0.777 [0.777] | 6.114 | 6.197 | 6.090 |
| | (0.059) | | | | |
| Direct treatment vs. control | -0.081 | 0.303 [0.528] | 6.114 | 6.197 | 6.090 |
| | (0.077) | | | | |
| Direct treatment vs. spillover | -0.008 | 0.906 [0.995] | 6.114 | 6.197 | 6.090 |
| | (0.070) | | | | |
| *Panel H. Requests economics-program updates* | | | | | |
| Spillover vs. control | -2.301 | 0.433 [0.693] | 27.848 | 25.051 | 27.000 |
| | (2.893) | | | | |
| Direct treatment vs. control | -0.409 | 0.890 [0.982] | 27.848 | 25.051 | 27.000 |
| | (2.932) | | | | |
| Direct treatment vs. spillover | 3.339 | 0.230 [0.368] | 27.848 | 25.051 | 27.000 |
| | (2.761) | | | | |

*Notes:* Each panel reports a follow-up measure of exposure, recall, discussion, or demand for economics-program updates. Activity reports, naming a female economist, any/partial recall, exact recall, and requests for updates are binary: effects are in percentage points and arm means are percentages. Discussion with family/friends/peers, discussion initiated by someone else, and ability to describe an activity topic are 0–10 response scales: their effects and standard errors are divided by the outcome-specific control-group standard deviation, including for the within-school contrast, whereas their arm means remain on the original 0–10 scale. Any/partial recall includes partial responses; exact recall includes only affirmative responses. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No additional covariates are included. The $p$-value tests the displayed assignment effect; bracketed $q$-values use the Benjamini–Hochberg adjustment across all eight outcomes separately for each contrast, not separately by panel. The sample consists of final-year baseline respondents with the relevant follow-up measure observed. These post-assignment reports describe possible diffusion channels; they do not by themselves identify mediation or the effect of actual discussion. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S14: Exposure and Recall by Gender

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Outcome family: Follow-up exposure and recall; Outcome: Reports a female-economist activity* | | | | |
| Spillover vs. control | -0.209 | -2.226 | -2.017 | 0.612 [0.803] |
| | (3.256) | (3.151) | (3.926) | |
| Treated-school assignment | -1.958 | -2.227 | -0.269 | 0.935 [0.942] |
| | (3.092) | (2.279) | (3.273) | |
| Direct treatment vs. control | -3.557 | -2.363 | 1.194 | 0.806 [0.819] |
| | (4.205) | (3.633) | (4.825) | |
| Direct treatment vs. spillover | -0.667 | 1.156 | 1.823 | 0.701 [0.926] |
| | (4.490) | (4.823) | (4.734) | |
| *Panel B. Outcome family: Follow-up exposure and recall; Outcome: Reports activity, among yes/no responses* | | | | |
| Spillover vs. control | -0.664 | 0.687 | 1.351 | 0.586 [0.803] |
| | (2.359) | (1.914) | (2.449) | |
| Treated-school assignment | -1.129 | 0.312 | 1.441 | 0.550 [0.725] |
| | (1.751) | (1.473) | (2.380) | |
| Direct treatment vs. control | -1.612 | -0.183 | 1.429 | 0.672 [0.819] |
| | (2.188) | (2.261) | (3.342) | |
| Direct treatment vs. spillover | 1.243 | 0.669 | -0.574 | 0.869 [0.926] |
| | (3.405) | (3.349) | (3.466) | |
| *Panel C. Outcome family: Follow-up exposure and recall; Outcome: Discussed economics with family, friends, or peers* | | | | |
| Spillover vs. control | 0.022 | 0.390 | 0.368 | 0.259 [0.611] |
| | (0.285) | (0.241) | (0.319) | |
| Treated-school assignment | -0.099 | 0.448* | 0.546* | 0.064 [0.330] |
| | (0.188) | (0.220) | (0.283) | |
| Direct treatment vs. control | -0.201 | 0.507 | 0.709** | 0.043 [0.249] |
| | (0.230) | (0.320) | (0.334) | |
| Direct treatment vs. spillover | -0.023 | 0.289 | 0.313 | 0.334 [0.926] |
| | (0.312) | (0.275) | (0.321) | |
| *Panel D. Outcome family: Follow-up exposure and recall; Outcome: Can name a female economist* | | | | |
| Spillover vs. control | 1.343 | 6.420* | 5.077 | 0.356 [0.721] |
| | (4.739) | (3.629) | (5.410) | |
| Treated-school assignment | 3.685 | 9.970*** | 6.286 | 0.125 [0.470] |
| | (3.776) | (3.537) | (3.971) | |
| Direct treatment vs. control | 6.070 | 14.320*** | 8.251* | 0.076 [0.281] |
| | (3.917) | (4.507) | (4.470) | |
| Direct treatment vs. spillover | 5.216 | 5.815 | 0.599 | 0.918 [0.926] |
| | (4.312) | (4.147) | (5.812) | |

*Continued on next page*

Table S14 continued

| Assignment contrast | Men | Women | Women - men | $p$-value [wild $q$-value] |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel E. Outcome family: Follow-up exposure and recall; Outcome: Someone discussed studying economics with student* | | | | |
| Spillover vs. control | -0.052 | -0.145 | -0.093 | 0.818 [0.831] |
| | (0.303) | (0.324) | (0.400) | |
| Treated-school assignment | -0.213 | 0.045 | 0.257 | 0.487 [0.725] |
| | (0.230) | (0.303) | (0.365) | |
| Direct treatment vs. control | -0.343 | 0.258 | 0.601 | 0.132 [0.355] |
| | (0.220) | (0.333) | (0.387) | |
| Direct treatment vs. spillover | 0.039 | 0.563** | 0.524 | 0.175 [0.926] |
| | (0.309) | (0.250) | (0.384) | |

*Notes:* The outcomes are follow-up reports of activity exposure, naming a female economist, and economics-related discussion. Panels A, B, and D use binary indicators multiplied by 100, so effects are in percentage points. Panel A codes an affirmative activity report against all other nonmissing responses; Panel B restricts this comparison to yes/no answers. Panels C and E are 0–10 discussion scales, and their coefficients are in original scale points, not standard deviations. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No additional covariates are included. The reported $p$-value tests the women-minus-men difference using conventional clustered inference. Bracketed $q$-values apply the Benjamini–Hochberg adjustment to symmetric wild-cluster-bootstrap tests of the same interaction, using 5,000 Rademacher replications and the clustering level specified above. Adjustment is separate for each assignment contrast within the original nine follow-up exposure, recall, discussion, and update-request outcomes, jointly across the two presentation tables. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These post-assignment responses describe exposure, reactions, or potential channels; they do not identify causal mediation. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S15: Diffusion and Follow-Up Information Requests by Gender

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Outcome family: Follow-up exposure and recall; Outcome: Any or partial recall of visiting economist* | | | | |
| Spillover vs. control | 12.487 | 19.385 | 6.898 | 0.208 [0.611] |
| | (13.525) | (14.621) | (5.342) | |
| Treated-school assignment | 32.407** | 34.934** | 2.527 | 0.489 [0.725] |
| | (12.332) | (12.622) | (3.598) | |
| Direct treatment vs. control | 51.934*** | 55.087*** | 3.153 | 0.360 [0.490] |
| | (12.186) | (12.084) | (3.387) | |
| Direct treatment vs. spillover | 38.002*** | 33.043*** | -4.959 | 0.435 [0.926] |
| | (7.467) | (7.224) | (6.325) | |
| *Panel B. Outcome family: Follow-up exposure and recall; Outcome: Exact recall of visiting economist* | | | | |
| Spillover vs. control | 11.725 | 17.145 | 5.419 | 0.273 [0.611] |
| | (13.381) | (14.059) | (4.843) | |
| Treated-school assignment | 33.299** | 33.880** | 0.581 | 0.868 [0.942] |
| | (12.048) | (12.273) | (3.451) | |
| Direct treatment vs. control | 54.437*** | 55.579*** | 1.142 | 0.753 [0.819] |
| | (11.834) | (11.965) | (3.594) | |
| Direct treatment vs. spillover | 40.929*** | 35.755*** | -5.174 | 0.413 [0.926] |
| | (7.927) | (7.557) | (6.289) | |
| *Panel C. Outcome family: Follow-up exposure and recall; Outcome: Can describe at least one topic from activity* | | | | |
| Spillover vs. control | -0.396 | 0.237 | 0.633** | 0.036 [0.409] |
| | (0.236) | (0.210) | (0.286) | |
| Treated-school assignment | -0.498** | 0.161 | 0.658*** | 0.005 [0.142] |
| | (0.181) | (0.201) | (0.214) | |
| Direct treatment vs. control | -0.597** | 0.062 | 0.660** | 0.033 [0.249] |
| | (0.259) | (0.265) | (0.294) | |
| Direct treatment vs. spillover | 0.023 | -0.060 | -0.084 | 0.806 [0.926] |
| | (0.298) | (0.238) | (0.339) | |

*Continued on next page*

Table S15 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel D. Outcome family: Follow-up exposure and recall; Outcome: Requests economics-program updates* | | | | |
| Spillover vs. control | -3.080 | -1.405 | 1.675 | 0.733 [0.830] |
| | (4.671) | (2.943) | (4.852) | |
| Treated-school assignment | -3.855 | 0.727 | 4.582 | 0.317 [0.725] |
| | (4.055) | (2.914) | (4.492) | |
| Direct treatment vs. control | -4.387 | 3.149 | 7.537 | 0.195 [0.370] |
| | (4.386) | (3.771) | (5.668) | |
| Direct treatment vs. spillover | 1.967 | 4.641 | 2.674 | 0.651 [0.926] |
| | (4.423) | (3.783) | (5.883) | |

*Notes:* Panels A and B distinguish any/partial recall from exact recall of the visiting economist; Panel C measures ability to describe an activity topic; Panel D indicates a request for economics-program updates. The recall and request indicators are multiplied by 100 and reported in percentage points. The topic response is measured on a 0–10 scale, and its coefficient and standard error remain in original scale points. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No additional covariates are included. The reported $p$-value tests the women-minus-men difference using conventional clustered inference. Bracketed $q$-values apply the Benjamini–Hochberg adjustment to symmetric wild-cluster-bootstrap tests of the same interaction, using 5,000 Rademacher replications and the clustering level specified above. Adjustment is separate for each assignment contrast within the original nine follow-up exposure, recall, discussion, and update-request outcomes, jointly across the two presentation tables. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These post-assignment responses describe exposure, reactions, or potential channels; they do not identify causal mediation. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S16: Information and Perceived-Barriers Mechanisms

| Outcome | Assignment contrast | Effect | $p$-value [$q$-value] | Control mean | Spillover mean | Direct-treatment mean |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A. Outcome family: barriers* | | | | | | |
| Barrier selected: cost | Treated-school assignment | -3.517<br>(2.612) | 0.189 [0.505] | 22.348 | 18.421 | 19.178 |
| Barrier selected: distance | Treated-school assignment | -1.272<br>(2.054) | 0.541 [0.718] | 19.639 | 14.170 | 19.765 |
| Barrier selected: family expectations | Treated-school assignment | -0.459<br>(0.939) | 0.629 [0.718] | 5.530 | 5.061 | 5.088 |
| Barrier selected: cannot imagine an economics career | Treated-school assignment | 5.706**<br>(2.135) | 0.013 [0.101] | 43.002 | 49.595 | 43.444 |
| Barrier selected: does not know what economists do | Treated-school assignment | -1.952<br>(2.056) | 0.351 [0.702] | 29.007 | 30.162 | 24.658 |
| Barrier selected: low mathematics confidence | Treated-school assignment | -1.044<br>(3.126) | 0.741 [0.741] | 31.038 | 27.733 | 29.746 |
| Barrier selected: other | Treated-school assignment | -0.289<br>(0.564) | 0.612 [0.718] | 3.047 | 3.036 | 3.327 |
| Barrier selected: prefers another field | Treated-school assignment | 3.958<br>(2.775) | 0.165 [0.505] | 42.438 | 48.381 | 48.728 |

*Continued on next page*

Table S16 continued

| Outcome | Assignment contrast | Effect | $p$-value [$q$-value] | Control mean | Spillover mean | Direct-treatment mean |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel B. Outcome family: information* | | | | | | |
| Information requested: financial aid or scholarships | Treated-school assignment | -0.448 (3.225) | 0.891 [0.931] | 29.886 | 29.960 | 26.719 |
| Information requested: admission dates | Treated-school assignment | -0.566 (1.222) | 0.647 [0.931] | 11.023 | 9.312 | 10.020 |
| Information requested: economics labor market | Treated-school assignment | 3.487 (3.223) | 0.289 [0.931] | 27.841 | 28.340 | 31.041 |
| Information requested: none | Treated-school assignment | -0.482 (2.928) | 0.871 [0.931] | 35.000 | 38.664 | 35.167 |
| Information requested: economics-program brochures | Treated-school assignment | 2.292 (2.264) | 0.320 [0.931] | 20.568 | 21.053 | 19.253 |
| Information requested: economics-career talks | Treated-school assignment | 0.284 (3.243) | 0.931 [0.931] | 27.159 | 26.518 | 23.772 |
| Information requested: student/alumna testimonials | Treated-school assignment | 0.227 (1.960) | 0.909 [0.931] | 24.886 | 25.506 | 20.236 |

*Notes:* Panel A reports indicators for each perceived barrier to studying economics; Panel B reports indicators for each type of information requested at follow-up. Students could select multiple options, so the percentages need not sum to 100. Each indicator is multiplied by 100: effects and standard errors are in percentage points, whereas the three arm means are unadjusted percentages. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Each regression includes randomization-block fixed effects and uses school-clustered standard errors, without additional covariates. The $p$-value tests the treated-school effect; bracketed $q$-values use the Benjamini–Hochberg false-discovery-rate adjustment separately across the eight barrier items and the seven information items. The sample comprises final-year baseline respondents with a nonmissing answer to the relevant follow-up question; a missing answer is not coded as an unselected option. These exercises document deviations from the pre-analysis plan (PAP): barriers were collected as an expanded multi-select question, and information options were also expanded. They measure reported constraints and information demand, not identified causal mediation. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S17: Perceived Barriers by Gender

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Outcome family: barriers; Outcome: Barrier selected: cost* | | | | |
| Spillover vs. control | -2.738 | -3.927 | -1.189 | 0.836 [0.861] |
| | (4.773) | (3.115) | (5.704) | |
| Treated-school assignment | -3.369 | -3.454 | -0.086 | 0.987 [0.988] |
| | (4.382) | (2.688) | (5.123) | |
| Direct treatment vs. control | -3.890 | -2.948 | 0.942 | 0.853 [0.854] |
| | (4.477) | (3.015) | (5.038) | |
| Direct treatment vs. spillover | 0.265 | 1.973 | 1.708 | 0.770 [0.937] |
| | (4.463) | (3.586) | (5.831) | |
| *Panel B. Outcome family: barriers; Outcome: Barrier selected: distance* | | | | |
| Spillover vs. control | -3.324 | -4.280 | -0.956 | 0.838 [0.861] |
| | (4.439) | (2.581) | (4.623) | |
| Treated-school assignment | -3.813 | 0.174 | 3.987 | 0.338 [0.751] |
| | (3.308) | (2.467) | (4.084) | |
| Direct treatment vs. control | -3.853 | 5.396* | 9.249* | 0.081 [0.279] |
| | (3.967) | (2.999) | (5.108) | |
| Direct treatment vs. spillover | -1.711 | 8.561** | 10.272* | 0.079 [0.672] |
| | (4.137) | (3.451) | (5.772) | |
| *Panel C. Outcome family: barriers; Outcome: Barrier selected: family expectations* | | | | |
| Spillover vs. control | 1.679 | -1.557 | -3.236 | 0.193 [0.550] |
| | (1.916) | (1.294) | (2.425) | |
| Treated-school assignment | 0.817 | -1.242 | -2.059 | 0.348 [0.751] |
| | (1.739) | (1.128) | (2.156) | |
| Direct treatment vs. control | 0.075 | -0.942 | -1.017 | 0.657 [0.854] |
| | (2.144) | (1.338) | (2.267) | |
| Direct treatment vs. spillover | -1.980 | 0.503 | 2.483 | 0.357 [0.933] |
| | (2.431) | (1.403) | (2.683) | |
| *Panel D. Outcome family: barriers; Outcome: Barrier selected: cannot imagine an economics career* | | | | |
| Spillover vs. control | 6.315 | 7.310** | 0.995 | 0.858 [0.861] |
| | (3.744) | (3.161) | (5.513) | |
| Treated-school assignment | 6.724** | 4.143 | -2.581 | 0.569 [0.770] |
| | (2.671) | (3.336) | (4.475) | |
| Direct treatment vs. control | 6.808* | 0.436 | -6.372 | 0.303 [0.615] |
| | (3.354) | (4.869) | (6.063) | |
| Direct treatment vs. spillover | -0.103 | -6.602 | -6.499 | 0.353 [0.933] |
| | (4.816) | (4.584) | (6.960) | |

*Continued on next page*

Table S17 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel E. Outcome family: barriers; Outcome: Barrier selected: does not know what economists do* | | | | |
| Spillover vs. control | 7.044* | -3.455 | -10.499* | 0.078 [0.424] |
| | (3.880) | (3.469) | (5.729) | |
| Treated-school assignment | 2.931 | -5.741* | -8.672** | 0.038 [0.186] |
| | (2.563) | (3.072) | (3.967) | |
| Direct treatment vs. control | -0.948 | -8.781** | -7.833* | 0.054 [0.279] |
| | (3.228) | (3.971) | (3.886) | |
| Direct treatment vs. spillover | -6.752 | -4.471 | 2.281 | 0.705 [0.937] |
| | (4.889) | (4.110) | (5.998) | |
| *Panel F. Outcome family: barriers; Outcome: Barrier selected: low mathematics confidence* | | | | |
| Spillover vs. control | 3.010 | -6.934* | -9.944** | 0.019 [0.222] |
| | (2.644) | (4.028) | (3.990) | |
| Treated-school assignment | 4.117 | -5.471 | -9.588** | 0.020 [0.186] |
| | (2.737) | (4.140) | (3.886) | |
| Direct treatment vs. control | 5.236 | -3.651 | -8.887* | 0.067 [0.279] |
| | (3.562) | (4.835) | (4.661) | |
| Direct treatment vs. spillover | 1.633 | 2.784 | 1.151 | 0.818 [0.937] |
| | (4.138) | (3.617) | (4.987) | |
| *Panel G. Outcome family: barriers; Outcome: Barrier selected: other* | | | | |
| Spillover vs. control | -0.626 | 0.111 | 0.737 | 0.720 [0.861] |
| | (1.583) | (0.880) | (2.036) | |
| Treated-school assignment | -0.607 | 0.096 | 0.703 | 0.713 [0.842] |
| | (1.343) | (0.892) | (1.889) | |
| Direct treatment vs. control | -0.591 | 0.080 | 0.671 | 0.807 [0.854] |
| | (2.089) | (1.205) | (2.722) | |
| Direct treatment vs. spillover | -0.362 | -0.172 | 0.190 | 0.943 [0.943] |
| | (2.263) | (1.178) | (2.631) | |

*Continued on next page*

Table S17 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel H. Outcome family: barriers; Outcome: Barrier selected: prefers another field* | | | | |
| Spillover vs. control | 6.527 | 1.997 | -4.530 | 0.375 [0.768] |
| | (4.164) | (3.966) | (5.022) | |
| Treated-school assignment | 5.690 | 2.376 | -3.314 | 0.465 [0.751] |
| | (3.693) | (3.267) | (4.471) | |
| Direct treatment vs. control | 4.976 | 2.755 | -2.221 | 0.722 [0.854] |
| | (4.093) | (4.090) | (6.181) | |
| Direct treatment vs. spillover | -1.003 | 0.898 | 1.902 | 0.816 [0.937] |
| | (6.073) | (4.339) | (8.150) | |

*Notes:* Each outcome indicates selection of a particular barrier to studying economics in the follow-up multi-select question. Indicators are multiplied by 100; all effects and standard errors are in percentage points. Multiple barriers can be selected, so categories are not mutually exclusive. Missing answers are excluded, not coded as nonselection. The expanded multi-select instrument differs from the pre-analysis plan (PAP)'s barrier question. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No additional covariates are included. The reported $p$-value tests the women-minus-men difference using conventional clustered inference. Bracketed $q$-values apply the Benjamini–Hochberg adjustment to symmetric wild-cluster-bootstrap tests of the same interaction, using 5,000 Rademacher replications and the clustering level specified above. Adjustment is separate for each assignment contrast within the eight barrier outcomes. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These post-assignment responses describe exposure, reactions, or potential channels; they do not identify causal mediation. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S18: Information Requests by Gender

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Outcome family: information; Outcome: Information requested: financial aid or scholarships* | | | | |
| Spillover vs. control | 0.470 | 2.061 | 1.591 | 0.718 [0.990] |
| | (4.262) | (4.654) | (4.364) | |
| Treated-school assignment | -1.571 | 1.035 | 2.607 | 0.564 [0.973] |
| | (3.824) | (3.868) | (4.457) | |
| Direct treatment vs. control | -3.484 | -0.349 | 3.134 | 0.547 [0.875] |
| | (4.109) | (4.559) | (5.134) | |
| Direct treatment vs. spillover | -2.979 | -1.124 | 1.855 | 0.734 [0.951] |
| | (4.224) | (4.007) | (5.446) | |
| *Panel B. Outcome family: information; Outcome: Information requested: admission dates* | | | | |
| Spillover vs. control | -0.917 | -0.703 | 0.214 | 0.950 [0.990] |
| | (2.427) | (1.923) | (3.409) | |
| Treated-school assignment | -0.056 | -0.756 | -0.700 | 0.808 [0.973] |
| | (2.217) | (1.521) | (2.856) | |
| Direct treatment vs. control | 0.709 | -0.748 | -1.457 | 0.643 [0.875] |
| | (2.351) | (1.751) | (3.108) | |
| Direct treatment vs. spillover | 1.285 | 0.447 | -0.839 | 0.853 [0.951] |
| | (3.272) | (2.486) | (4.517) | |
| *Panel C. Outcome family: information; Outcome: Information requested: economics labor market* | | | | |
| Spillover vs. control | 0.451 | 3.370 | 2.919 | 0.635 [0.990] |
| | (4.127) | (4.752) | (6.082) | |
| Treated-school assignment | 0.341 | 6.220 | 5.880 | 0.216 [0.723] |
| | (3.853) | (4.103) | (4.640) | |
| Direct treatment vs. control | 0.489 | 9.591** | 9.101* | 0.078 [0.644] |
| | (4.115) | (4.626) | (4.963) | |
| Direct treatment vs. spillover | 2.166 | 7.596* | 5.430 | 0.361 [0.951] |
| | (4.101) | (3.888) | (5.906) | |
| *Panel D. Outcome family: information; Outcome: Information requested: none* | | | | |
| Spillover vs. control | 3.206 | -0.204 | -3.410 | 0.528 [0.990] |
| | (3.797) | (4.411) | (5.331) | |
| Treated-school assignment | 1.938 | -2.915 | -4.853 | 0.292 [0.723] |
| | (3.771) | (3.656) | (4.515) | |
| Direct treatment vs. control | 0.570 | -6.234 | -6.804 | 0.182 [0.736] |
| | (4.669) | (4.417) | (4.961) | |
| Direct treatment vs. spillover | -5.588 | -6.007 | -0.419 | 0.949 [0.951] |
| | (5.361) | (4.026) | (6.526) | |
| *Panel E. Outcome family: information; Outcome: Information requested: economics-program brochures* | | | | |
| Spillover vs. control | -0.003 | 4.870 | 4.873 | 0.316 [0.990] |
| | (3.207) | (3.469) | (4.772) | |

*Continued on next page*

Table S18 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| Treated-school assignment | -0.348 | 4.316 | 4.664 | 0.218 [0.723] |
| | (2.706) | (3.105) | (3.702) | |
| Direct treatment vs. control | -0.703 | 3.632 | 4.335 | 0.377 [0.875] |
| | (3.748) | (3.287) | (4.822) | |
| Direct treatment vs. spillover | -0.114 | -1.483 | -1.368 | 0.805 [0.951] |
| | (4.119) | (2.901) | (5.522) | |
| *Panel F. Outcome family: information; Outcome: Information requested: economics-career talks* | | | | |
| Spillover vs. control | 1.180 | 1.258 | 0.078 | 0.988 [0.990] |
| | (4.806) | (3.102) | (5.307) | |
| Treated-school assignment | 0.158 | 0.542 | 0.384 | 0.933 [0.973] |
| | (4.626) | (3.364) | (4.549) | |
| Direct treatment vs. control | -0.817 | -0.391 | 0.426 | 0.936 [0.942] |
| | (4.929) | (4.551) | (5.254) | |
| Direct treatment vs. spillover | -0.884 | -1.721 | -0.837 | 0.876 [0.951] |
| | (4.022) | (3.402) | (5.365) | |
| *Panel G. Outcome family: information; Outcome: Information requested: student/alumna testimonials* | | | | |
| Spillover vs. control | 1.493 | 3.097 | 1.604 | 0.742 [0.990] |
| | (3.451) | (3.064) | (4.819) | |
| Treated-school assignment | 0.384 | 0.552 | 0.169 | 0.970 [0.973] |
| | (3.326) | (2.716) | (4.440) | |
| Direct treatment vs. control | -0.828 | -2.557 | -1.729 | 0.730 [0.875] |
| | (4.287) | (2.893) | (4.964) | |
| Direct treatment vs. spillover | -2.734 | -4.734* | -2.000 | 0.684 [0.951] |
| | (3.639) | (2.829) | (4.895) | |

*Notes:* Each outcome indicates selection of one follow-up information option: financial support, admission dates, labor-market information, brochures, career talks, testimonials, or no information. Indicators are multiplied by 100, so coefficients and standard errors are in percentage points. Students could select multiple options; a missing answer is not a nonselection. These options were expanded relative to the pre-analysis plan (PAP). Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No additional covariates are included. The reported $p$-value tests the women-minus-men difference using conventional clustered inference. Bracketed $q$-values apply the Benjamini–Hochberg adjustment to symmetric wild-cluster-bootstrap tests of the same interaction, using 5,000 Rademacher replications and the clustering level specified above. Adjustment is separate for each assignment contrast within the seven information-option outcomes. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These post-assignment responses describe exposure, reactions, or potential channels; they do not identify causal mediation. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S19: Treatment Effects on Academic Expectations and Quiz Performance

| Outcome | Effect (1) | Effect (SD) (2) | $p$-value [$q$-value] (3) | Control mean (4) | Control SD (5) |
|---|---|---|---|---|---|
| Subjective probability of scoring above the classroom median (0-100) | 0.912 | 0.031 | 0.683 [0.768] | 55.297 | 29.403 |
| | (2.206) | (0.075) | | | |
| Expected percentile in the classroom (0-100) | 1.761 | 0.064 | 0.430 [0.710] | 57.793 | 27.659 |
| | (2.198) | (0.079) | | | |
| Quiz score above sample median | 0.030 | 0.059 | 0.513 [0.710] | 0.444 | 0.497 |
| | (0.044) | (0.090) | | | |
| Realized percentile in brief quiz | 3.121 | 0.110 | 0.230 [0.691] | 47.910 | 28.343 |
| | (2.544) | (0.090) | | | |
| Brief quantitative quiz score (0-6; exploratory) | 0.256 | 0.123 | 0.173 [0.691] | 3.646 | 2.092 |
| | (0.183) | (0.088) | | | |

*Notes:* Students were asked to imagine a mathematics/economics exam taken by everyone in their classroom and to report their expected percentile and probability of scoring above the classroom median, both on a 0–100 scale. The brief quantitative quiz contains six questions, and its score is the number of correct answers, from 0 to 6. Realized percentiles and the above-median indicator use the pooled distribution of quiz scores among final-year baseline respondents with observed scores. The indicator is coded 0/1. Because the subjective classroom exam and the sample-wide quiz benchmark differ, these outcomes are reported separately and do not establish the accuracy of students' expectations. Effect reports original outcome units; Effect (SD) divides the estimate and standard error by the outcome-specific control-group standard deviation. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Regressions include randomization-block fixed effects, no additional covariates, and school-clustered standard errors. Bracketed $q$-values retain the Benjamini–Hochberg adjustment across the original nine conventional effect tests; the adjustment is not recomputed for the five outcomes displayed. The sample includes final-year baseline respondents with the relevant follow-up measures. Control mean and Control SD report unadjusted control moments in original units. These outcomes were not prespecified in the pre-analysis plan (PAP). The data come from the study's baseline and subsequent student surveys. Standard errors 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 S20: Treatment Effects on Academic Expectations by Gender

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Subjective probability of scoring above the classroom median (0–100)* | | | | |
| Spillover vs. control | -1.225 | 4.051 | 5.276* | 0.059 [0.437] |
| | (2.696) | (2.735) | (2.678) | |
| Treated-school assignment | -3.459 | 5.285** | 8.744*** | 0.003 [0.022] |
| | (2.407) | (2.277) | (2.664) | |
| Direct treatment vs. control | -5.354* | 6.500** | 11.854*** | 0.002 [0.010] |
| | (2.784) | (2.610) | (3.405) | |
| Direct treatment vs. spillover | -2.799 | 3.101 | 5.900 | 0.110 [0.337] |
| | (3.106) | (2.966) | (3.656) | |
| *Panel B. Expected percentile in the hypothetical classroom exam (0–100)* | | | | |
| Spillover vs. control | -0.314 | 3.985 | 4.298 | 0.161 [0.565] |
| | (3.129) | (2.594) | (2.980) | |
| Treated-school assignment | -2.090 | 5.722*** | 7.812*** | $<$ 0.001 [0.011] |
| | (2.267) | (2.038) | (2.012) | |
| Direct treatment vs. control | -3.533 | 7.555*** | 11.088*** | $<$ 0.001 [0.010] |
| | (2.467) | (2.162) | (2.539) | |
| Direct treatment vs. spillover | -1.824 | 4.332 | 6.155* | 0.055 [0.337] |
| | (2.714) | (2.943) | (3.158) | |

*Notes:* Panel A reports the subjective probability of scoring above the classroom median; Panel B reports expected percentile rank in a hypothetical mathematics/economics exam taken with classroom peers. Both responses range from 0 to 100, with higher values indicating a more favorable assessment of relative academic performance. Coefficients and standard errors are in original scale points. These are reported expectations; they do not measure forecast accuracy against the separate sample-wide quiz benchmark. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No additional covariates are included. The reported $p$-value tests the women-minus-men difference using conventional clustered inference. Bracketed $q$-values apply the Benjamini–Hochberg adjustment to symmetric wild-cluster-bootstrap tests of the same interaction, using 5,000 Rademacher replications and the clustering level specified above. Adjustment is separate for each assignment contrast within the original nine expectation, performance, and calibration outcomes, including archived measures omitted from the displayed tables. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These expectation outcomes were not prespecified in the pre-analysis plan (PAP). The data come from the study's baseline and subsequent student surveys. Standard errors 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 S21: Realized Quantitative Performance by Gender

| Assignment contrast | Men<br>(1) | Women<br>(2) | Women - men<br>(3) | $p$-value [wild $q$-value]<br>(4) |
|---|---|---|---|---|
| *Panel A. Outcome family: calibration; Outcome: Quiz score above sample median* | | | | |
| Spillover vs. control | 0.042 | 0.055 | 0.013 | 0.810 [0.832] |
| | (0.064) | (0.051) | (0.054) | |
| Treated-school assignment | 0.021 | 0.044 | 0.024 | 0.639 [0.654] |
| | (0.058) | (0.041) | (0.050) | |
| Direct treatment vs. control | 0.000 | 0.029 | 0.029 | 0.629 [0.727] |
| | (0.062) | (0.056) | (0.060) | |
| Direct treatment vs. spillover | -0.013 | -0.007 | 0.006 | 0.924 [0.921] |
| | (0.064) | (0.057) | (0.058) | |
| *Panel B. Outcome family: calibration; Outcome: Realized percentile in brief quiz* | | | | |
| Spillover vs. control | 3.162 | 5.596* | 2.434 | 0.442 [0.605] |
| | (3.904) | (2.757) | (3.122) | |
| Treated-school assignment | 2.154 | 4.394* | 2.240 | 0.410 [0.485] |
| | (3.323) | (2.175) | (2.679) | |
| Direct treatment vs. control | 1.129 | 2.881 | 1.753 | 0.564 [0.727] |
| | (3.402) | (3.016) | (3.004) | |
| Direct treatment vs. spillover | -1.126 | -2.173 | -1.047 | 0.733 [0.921] |
| | (3.472) | (3.326) | (3.055) | |

*Continued on next page*

Table S21 continued

| Assignment contrast | Men | Women | Women - men | $p$-value [wild $q$-value] |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel C. Outcome family: calibration; Outcome: Brief quantitative quiz score (0-6; exploratory)* | | | | |
| Spillover vs. control | 0.260 | 0.403* | 0.143 | 0.509 [0.605] |
| | (0.276) | (0.203) | (0.214) | |
| Treated-school assignment | 0.187 | 0.345** | 0.159 | 0.391 [0.485] |
| | (0.230) | (0.161) | (0.182) | |
| Direct treatment vs. control | 0.115 | 0.271 | 0.156 | 0.470 [0.727] |
| | (0.230) | (0.222) | (0.213) | |
| Direct treatment vs. spillover | -0.071 | -0.103 | -0.031 | 0.890 [0.921] |
| | (0.241) | (0.243) | (0.227) | |

*Notes:* Panel A uses an indicator that the brief quiz score exceeds the median among final-semester baseline respondents with observed quiz scores. This indicator remains coded 0/1: its coefficient is a probability difference, and multiplying it by 100 converts it to percentage points. Panel B uses realized percentile rank on a 0–100 scale, and Panel C the number of correct quiz answers on a 0–6 scale. Estimates and standard errors remain in these original units; none is standardized. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No additional covariates are included. The reported $p$-value tests the women-minus-men difference using conventional clustered inference. Bracketed $q$-values apply the Benjamini–Hochberg adjustment to symmetric wild-cluster-bootstrap tests of the same interaction, using 5,000 Rademacher replications and the clustering level specified above. Adjustment is separate for each assignment contrast within the original nine expectation, performance, and calibration outcomes, including archived measures omitted from the displayed tables. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These calibration and performance analyses were not prespecified in the pre-analysis plan (PAP). The data come from the study's baseline and subsequent student surveys. Standard errors 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 S22: Treatment Effects on Post-Intervention Mathematics Self-Assessment

| Contrast | Effect (SD) (1) | $p$-value [$q$-value] (2) | Control mean (3) | Control SD (4) |
|---|---|---|---|---|
| Spillover vs. control | -0.013 | 0.886 [0.990] | 6.458 | 2.719 |
| | (0.088) | | | |
| Treated-school assignment | -0.063 | 0.490 [0.623] | 6.458 | 2.719 |
| | (0.089) | | | |
| Direct treatment vs. control | -0.115 | 0.283 [0.494] | 6.458 | 2.719 |
| | (0.105) | | | |
| Direct treatment vs. spillover | -0.099 | 0.148 [0.344] | 6.662 | 2.755 |
| | (0.067) | | | |

*Notes:* The outcome is post-intervention mathematics self-assessment, measured on a 0-10 scale; this is a self-assessment, not a preference measure. The sample comprises final-semester students in the baseline sample with an observed response to this post-assignment measure. Treated-school assignment compares all students in treated schools, including directly treated and spillover classrooms, with students in control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with classrooms in control schools. These cross-school specifications include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, with school fixed effects and classroom-clustered standard errors. Effects and their standard errors are divided by the control-group standard deviation for cross-school comparisons and by the spillover-group standard deviation for the within-school comparison. Control mean and Control SD report these reference groups' unstandardized moments; thus, the last row reports the spillover mean and standard deviation. Bracketed $q$-values retain the original Benjamini–Hochberg adjustment across all fourteen post-assignment measures, separately for each contrast; the other thirteen measures are retained in the replication results but are not displayed as outcome tables in the Online Appendix. The editorial removal does not redefine the testing family. These outcomes were not prespecified in the PAP and are exploratory. Standard errors 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 S23: Treatment Effects on Mathematics Self-Assessment by Gender

| Assignment contrast | Men | Women | Women - men | $p$-value [wild $q$-value] |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Spillover vs. control | -0.201 | 0.153 | 0.355 | 0.142 [0.430] |
| | (0.234) | (0.287) | (0.234) | |
| Treated-school assignment | -0.403 | 0.039 | 0.443* | 0.054 [0.321] |
| | (0.280) | (0.246) | (0.219) | |
| Direct treatment vs. control | -0.597 | -0.093 | 0.504* | 0.052 [0.437] |
| | (0.365) | (0.238) | (0.248) | |
| Direct treatment vs. spillover | -0.408* | -0.236 | 0.172 | 0.538 [0.991] |
| | (0.228) | (0.226) | (0.278) | |

*Notes:* The dependent variable is self-assessed mathematical ability on a 0–10 response scale, with higher values indicating greater perceived ability. Coefficients and standard errors are in original scale points. The item measures perceived competence and is distinct from performance on an objective mathematics test. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No additional covariates are included. The reported $p$-value tests the women-minus-men difference using conventional clustered inference. Bracketed $q$-values apply the Benjamini–Hochberg adjustment to symmetric wild-cluster-bootstrap tests of the same interaction, using 5,000 Rademacher replications and the clustering level specified above. Adjustment is separate for each assignment contrast within the original fourteen post-assignment measures; the other thirteen measures remain in the adjustment only and are not displayed as outcomes. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These exploratory comparisons test whether the intervention changes perceived mathematical competence differently by student gender. The measure comes from the study's post-intervention self-assessment item. Standard errors 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 S24: Exploratory Survey Mediation with Beliefs about Economics as Mediator

| Contrast | Path $a$ | Path $b$ | Direct | Indirect | $p$-value [$q$-value], indirect |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Panel A. Outcome: Economics application: broad gateway* | | | | | |
| Treated-school assignment | 0.341 | 0.249 | 0.588 | 0.085 | 0.465 [0.896] |
| | (0.068) | (0.354) | (0.367) | (0.116) | |
| Direct treatment vs. control | 0.715 | 0.217 | 0.806 | 0.155 | 0.535 [0.967] |
| | (0.085) | (0.357) | (0.435) | (0.250) | |
| *Panel B. Outcome: Administrative applications* | | | | | |
| Treated-school assignment | 0.341 | 0.222 | 0.294 | 0.076 | 0.227 [0.455] |
| | (0.068) | (0.193) | (0.259) | (0.063) | |
| Direct treatment vs. control | 0.715 | 0.205 | 0.411 | 0.146 | 0.310 [0.619] |
| | (0.085) | (0.208) | (0.356) | (0.144) | |

*Notes:* The sample consists of women in the final year of four- or six-semester programs with an observed immediate mediator, application outcome, and baseline controls. Panel A uses the broad-gateway definition of economics applications (UANL Economics or the Tec ECSG gateway); Panel B uses the administrative measure (UANL Economics or Tec LEC). The immediate post-intervention beliefs-about-economics index is the proposed mediator and is expressed in baseline control-group standard deviations. Treated-school assignment compares all classrooms in treated schools with control schools. Direct treatment versus control compares direct-treatment classrooms with control schools, allowing separately for spillover classrooms. Path $a$ is the coefficient on the indicated treatment contrast in the mediator equation. Path $b$ is the mediator–application association conditional on assignment and controls. Direct is the assignment coefficient in that outcome equation; Indirect is the product $a \times b$. Application indicators are multiplied by 100: Direct and Indirect are in percentage points, and Path $b$ is in percentage points per standard deviation of the mediator. Both equations include randomization-block fixed effects. The mediator equation controls for its matched baseline index; the application equation controls for baseline age, GPA, application propensity, interest in economics, and self-assessed mathematics ability. Standard errors use the joint school-clustered covariance matrix; uncertainty for the product uses the delta method. Bracketed $q$-values retain the Benjamini–Hochberg adjustment across the four candidate mediators, separately for each application definition and treatment contrast. Each mediator is estimated separately, so the products are not additive contributions to the treatment effect. Random assignment does not identify the mediator–outcome association causally; that interpretation requires sequential ignorability conditional on the included controls. The reported products are exploratory. Standard errors 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 S25: Exploratory Survey Mediation with Identity and Self-Concept as Mediator

| Contrast | Path $a$ | Path $b$ | Direct | Indirect | $p$-value [$q$-value], indirect |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Panel A. Outcome: Economics application: broad gateway* | | | | | |
| Treated-school assignment | 0.177 | 0.450 | 0.613 | 0.080** | 0.025 [0.100] |
| | (0.026) | (0.190) | (0.363) | (0.036) | |
| Direct treatment vs. control | 0.299 | 0.431 | 0.841 | 0.129** | 0.019 [0.077] |
| | (0.043) | (0.180) | (0.445) | (0.055) | |
| *Panel B. Outcome: Administrative applications* | | | | | |
| Treated-school assignment | 0.177 | 0.436 | 0.312 | 0.077*** | 0.009 [0.036] |
| | (0.026) | (0.156) | (0.257) | (0.030) | |
| Direct treatment vs. control | 0.299 | 0.425 | 0.439 | 0.127*** | 0.009 [0.035] |
| | (0.043) | (0.155) | (0.336) | (0.048) | |

*Notes:* The sample consists of women in the final year of four- or six-semester programs with an observed immediate mediator, application outcome, and baseline controls. Panel A uses the broad-gateway definition of economics applications (UANL Economics or the Tec ECSG gateway); Panel B uses the administrative measure (UANL Economics or Tec LEC). The immediate post-intervention identity-and-self-concept index is the proposed mediator and is expressed in baseline control-group standard deviations. Treated-school assignment compares all classrooms in treated schools with control schools. Direct treatment versus control compares direct-treatment classrooms with control schools, allowing separately for spillover classrooms. Path $a$ is the coefficient on the indicated treatment contrast in the mediator equation. Path $b$ is the mediator–application association conditional on assignment and controls. Direct is the assignment coefficient in that outcome equation; Indirect is the product $a \times b$. Application indicators are multiplied by 100: Direct and Indirect are in percentage points, and Path $b$ is in percentage points per standard deviation of the mediator. Both equations include randomization-block fixed effects. The mediator equation controls for its matched baseline index; the application equation controls for baseline age, GPA, application propensity, interest in economics, and self-assessed mathematics ability. Standard errors use the joint school-clustered covariance matrix; uncertainty for the product uses the delta method. Bracketed $q$-values retain the Benjamini–Hochberg adjustment across the four candidate mediators, separately for each application definition and treatment contrast. Each mediator is estimated separately, so the products are not additive contributions to the treatment effect. Random assignment does not identify the mediator–outcome association causally; that interpretation requires sequential ignorability conditional on the included controls. The reported products are exploratory. Standard errors 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 S26: Exploratory Survey Mediation with Intentions and Aspirations as Mediator

| Contrast | Path $a$ | Path $b$ | Direct | Indirect | $p$-value [$q$-value], indirect |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Panel A. Outcome: Economics application: broad gateway* | | | | | |
| Treated-school assignment | -0.033 | 0.127 | 0.660 | -0.004 | 0.679 [0.896] |
| | (0.059) | (0.176) | (0.370) | (0.010) | |
| Direct treatment vs. control | -0.004 | 0.123 | 0.928 | -0.001 | 0.967 [0.967] |
| | (0.096) | (0.177) | (0.466) | (0.012) | |
| *Panel B. Outcome: Administrative applications* | | | | | |
| Treated-school assignment | -0.033 | 0.133 | 0.358 | -0.004 | 0.591 [0.788] |
| | (0.059) | (0.041) | (0.261) | (0.008) | |
| Direct treatment vs. control | -0.004 | 0.131 | 0.524 | -0.001 | 0.966 [0.966] |
| | (0.096) | (0.042) | (0.344) | (0.013) | |

*Notes:* The sample consists of women in the final year of four- or six-semester programs with an observed immediate mediator, application outcome, and baseline controls. Panel A uses the broad-gateway definition of economics applications (UANL Economics or the Tec ECSG gateway); Panel B uses the administrative measure (UANL Economics or Tec LEC). The immediate post-intervention intentions-and-aspirations index is the proposed mediator and is expressed in baseline control-group standard deviations. Treated-school assignment compares all classrooms in treated schools with control schools. Direct treatment versus control compares direct-treatment classrooms with control schools, allowing separately for spillover classrooms. Path $a$ is the coefficient on the indicated treatment contrast in the mediator equation. Path $b$ is the mediator–application association conditional on assignment and controls. Direct is the assignment coefficient in that outcome equation; Indirect is the product $a \times b$. Application indicators are multiplied by 100: Direct and Indirect are in percentage points, and Path $b$ is in percentage points per standard deviation of the mediator. Both equations include randomization-block fixed effects. The mediator equation controls for its matched baseline index; the application equation controls for baseline age, GPA, application propensity, interest in economics, and self-assessed mathematics ability. Standard errors use the joint school-clustered covariance matrix; uncertainty for the product uses the delta method. Bracketed $q$-values retain the Benjamini–Hochberg adjustment across the four candidate mediators, separately for each application definition and treatment contrast. Each mediator is estimated separately, so the products are not additive contributions to the treatment effect. Random assignment does not identify the mediator–outcome association causally; that interpretation requires sequential ignorability conditional on the included controls. The reported products are exploratory. Standard errors 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 S27: Exploratory Survey Mediation with Gender Stereotypes as Mediator

| Contrast | Path $a$ | Path $b$ | Direct | Indirect | $p$-value [$q$-value], indirect |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Panel A. Outcome: Economics application: broad gateway* | | | | | |
| Treated-school assignment | -0.007 | 0.114 | 0.531 | -0.001 | 0.896 [0.896] |
| | (0.054) | (0.205) | (0.329) | (0.007) | |
| Direct treatment vs. control | 0.014 | 0.098 | 1.069 | 0.001 | 0.815 [0.967] |
| | (0.060) | (0.206) | (0.473) | (0.006) | |
| *Panel B. Outcome: Administrative applications* | | | | | |
| Treated-school assignment | -0.007 | 0.203 | 0.233 | -0.002 | 0.893 [0.893] |
| | (0.054) | (0.166) | (0.294) | (0.011) | |
| Direct treatment vs. control | 0.014 | 0.194 | 0.534 | 0.003 | 0.808 [0.966] |
| | (0.060) | (0.166) | (0.393) | (0.011) | |

*Notes:* The sample consists of women in the final year of four- or six-semester programs with an observed immediate mediator, application outcome, and baseline controls. Panel A uses the broad-gateway definition of economics applications (UANL Economics or the Tec ECSG gateway); Panel B uses the administrative measure (UANL Economics or Tec LEC). The immediate post-intervention gender-stereotypes index is the proposed mediator and is expressed in baseline control-group standard deviations. Treated-school assignment compares all classrooms in treated schools with control schools. Direct treatment versus control compares direct-treatment classrooms with control schools, allowing separately for spillover classrooms. Path $a$ is the coefficient on the indicated treatment contrast in the mediator equation. Path $b$ is the mediator–application association conditional on assignment and controls. Direct is the assignment coefficient in that outcome equation; Indirect is the product $a \times b$. Application indicators are multiplied by 100: Direct and Indirect are in percentage points, and Path $b$ is in percentage points per standard deviation of the mediator. Both equations include randomization-block fixed effects. The mediator equation controls for its matched baseline index; the application equation controls for baseline age, GPA, application propensity, interest in economics, and self-assessed mathematics ability. Standard errors use the joint school-clustered covariance matrix; uncertainty for the product uses the delta method. Bracketed $q$-values retain the Benjamini–Hochberg adjustment across the four candidate mediators, separately for each application definition and treatment contrast. Each mediator is estimated separately, so the products are not additive contributions to the treatment effect. Random assignment does not identify the mediator–outcome association causally; that interpretation requires sequential ignorability conditional on the included controls. The reported products are exploratory. Standard errors 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 S28: Heterogeneity by Parental Education

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | No university-educated parent | 0.434 (0.474) | University-educated parent | 0.733 (0.650) | 0.299 (0.862) | 0.731 [0.950] |
| Administrative applications | No university-educated parent | 0.228 (0.413) | University-educated parent | 0.675 (0.513) | 0.447 (0.746) | 0.554 [0.950] |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | No university-educated parent | 0.164*** (0.057) | University-educated parent | 0.157** (0.075) | -0.007 (0.084) | 0.935 [0.995] |
| Identity and self-concept | No university-educated parent | 0.103* (0.054) | University-educated parent | -0.044 (0.068) | -0.147* (0.079) | 0.074 [0.583] |
| Intentions and aspirations | No university-educated parent | 0.061 (0.047) | University-educated parent | 0.015 (0.064) | -0.047 (0.074) | 0.537 [0.953] |
| Gender stereotypes | No university-educated parent | 0.088 (0.068) | University-educated parent | 0.037 (0.066) | -0.052 (0.083) | 0.541 [0.953] |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | No university-educated parent | 0.311*** (0.072) | University-educated parent | 0.284*** (0.064) | -0.027 (0.076) | 0.723 [0.869] |
| Identity and self-concept | No university-educated parent | 0.105*** (0.032) | University-educated parent | 0.091** (0.038) | -0.014 (0.047) | 0.764 [0.869] |
| Intentions and aspirations | No university-educated parent | -0.019 (0.076) | University-educated parent | -0.052 (0.075) | -0.033 (0.072) | 0.646 [0.869] |
| Gender stereotypes | No university-educated parent | 0.095** (0.040) | University-educated parent | 0.003 (0.055) | -0.092 (0.058) | 0.127 [0.869] |

*Notes:* The moderator is whether at least one parent has a university education. Panel A reports economics applications, Panel B follow-up survey outcomes, and Panel C immediate post-intervention survey outcomes. Each row is estimated from a pooled model containing treated-school assignment, the displayed subgroup indicator, and their interaction. Effect low and Effect high are subgroup-specific linear combinations; 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 expressed in baseline control-group standard deviations and oriented so that positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100; effects, standard errors, and control means are therefore reported in percentage points. The conventional $p$-value tests the interaction. The bracketed wild $q$-value applies a false-discovery-rate adjustment to null-imposed school wild-bootstrap $p$-values within the displayed moderator family. Standard errors 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 S29: Heterogeneity by School-Level Parental Education

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low school parental education | 0.595 (0.592) | High school parental education | 0.465 (0.747) | -0.130 (1.105) | 0.907 [0.950] |
| Administrative applications | Low school parental education | 0.661 (0.569) | High school parental education | -0.011 (0.476) | -0.671 (0.895) | 0.460 [0.950] |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low school parental education | 0.108 (0.071) | High school parental education | 0.259*** (0.076) | 0.152 (0.105) | 0.161 [0.953] |
| Identity and self-concept | Low school parental education | 0.081 (0.065) | High school parental education | 0.006 (0.075) | -0.075 (0.103) | 0.471 [0.953] |
| Intentions and aspirations | Low school parental education | 0.022 (0.041) | High school parental education | 0.017 (0.074) | -0.005 (0.092) | 0.958 [0.995] |
| Gender stereotypes | Low school parental education | 0.089 (0.078) | High school parental education | 0.061 (0.075) | -0.029 (0.111) | 0.798 [0.995] |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low school parental education | 0.320*** (0.091) | High school parental education | 0.288*** (0.079) | -0.032 (0.127) | 0.802 [0.869] |
| Identity and self-concept | Low school parental education | 0.090* (0.046) | High school parental education | 0.120*** (0.032) | 0.030 (0.063) | 0.644 [0.869] |
| Intentions and aspirations | Low school parental education | -0.089 (0.083) | High school parental education | 0.009 (0.065) | 0.098 (0.095) | 0.308 [0.869] |
| Gender stereotypes | Low school parental education | 0.017 (0.041) | High school parental education | 0.152*** (0.035) | 0.135*** (0.041) | 0.003 [0.157] |

*Notes:* The moderator is the school-level share of students with a university-educated parent; subgroup labels follow the prespecified split shown in the table. Panel A reports economics applications, Panel B follow-up survey outcomes, and Panel C immediate post-intervention survey outcomes. Each row is estimated from a pooled model containing treated-school assignment, the displayed subgroup indicator, and their interaction. Effect low and Effect high are subgroup-specific linear combinations; 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 expressed in baseline control-group standard deviations and oriented so that positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100; effects, standard errors, and control means are therefore reported in percentage points. The conventional $p$-value tests the interaction. The bracketed wild $q$-value applies a false-discovery-rate adjustment to null-imposed school wild-bootstrap $p$-values within the displayed moderator family. Standard errors 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 S30: Heterogeneity by Household Socioeconomic Status

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low SES assets | 0.372 | High SES assets | 0.621 | 0.249 | 0.699 [0.950] |
| | | (0.481) | | (0.458) | (0.637) | |
| Administrative applications | Low SES assets | 0.444 | High SES assets | 0.396 | -0.048 | 0.934 [0.950] |
| | | (0.482) | | (0.328) | (0.577) | |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low SES assets | 0.172** | High SES assets | 0.169** | -0.003 | 0.972 [0.995] |
| | | (0.064) | | (0.069) | (0.087) | |
| Identity and self-concept | Low SES assets | 0.098 | High SES assets | -0.015 | -0.112* | 0.057 [0.583] |
| | | (0.060) | | (0.049) | (0.056) | |
| Intentions and aspirations | Low SES assets | 0.011 | High SES assets | 0.010 | -0.001 | 0.995 [0.995] |
| | | (0.080) | | (0.049) | (0.100) | |
| Gender stereotypes | Low SES assets | -0.034 | High SES assets | 0.092 | 0.126 | 0.101 [0.583] |
| | | (0.074) | | (0.064) | (0.074) | |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low SES assets | 0.279** | High SES assets | 0.313*** | 0.034 | 0.701 [0.869] |
| | | (0.102) | | (0.053) | (0.087) | |
| Identity and self-concept | Low SES assets | 0.091 | High SES assets | 0.102*** | 0.012 | 0.865 [0.869] |
| | | (0.053) | | (0.033) | (0.068) | |
| Intentions and aspirations | Low SES assets | -0.025 | High SES assets | -0.040 | -0.016 | 0.869 [0.869] |
| | | (0.104) | | (0.061) | (0.095) | |
| Gender stereotypes | Low SES assets | 0.038 | High SES assets | 0.047 | 0.009 | 0.854 [0.869] |
| | | (0.044) | | (0.041) | (0.047) | |

*Notes:* The moderator is the baseline household socioeconomic-status index; subgroup labels follow the prespecified split shown in the table. Panel A reports economics applications, Panel B follow-up survey outcomes, and Panel C immediate post-intervention survey outcomes. Each row is estimated from a pooled model containing treated-school assignment, the displayed subgroup indicator, and their interaction. Effect low and Effect high are subgroup-specific linear combinations; 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 expressed in baseline control-group standard deviations and oriented so that positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100; effects, standard errors, and control means are therefore reported in percentage points. The conventional $p$-value tests the interaction. The bracketed wild $q$-value applies a false-discovery-rate adjustment to null-imposed school wild-bootstrap $p$-values within the displayed moderator family. Standard errors 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 S31: Heterogeneity by Baseline Academic Performance

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low GPA | 0.831** | High GPA | 0.257 | -0.574 | 0.471 [0.630] |
| | | (0.359) | | (0.647) | (0.785) | |
| Administrative applications | Low GPA | 0.520* | High GPA | 0.274 | -0.246 | 0.604 [0.630] |
| | | (0.272) | | (0.423) | (0.469) | |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low GPA | 0.175*** | High GPA | 0.159** | -0.015 | 0.866 [0.867] |
| | | (0.059) | | (0.076) | (0.089) | |
| Identity and self-concept | Low GPA | 0.102 | High GPA | -0.043 | -0.145* | 0.074 [0.301] |
| | | (0.060) | | (0.058) | (0.078) | |
| Intentions and aspirations | Low GPA | 0.032 | High GPA | -0.009 | -0.040 | 0.601 [0.805] |
| | | (0.056) | | (0.055) | (0.076) | |
| Gender stereotypes | Low GPA | 0.101 | High GPA | -0.003 | -0.104 | 0.285 [0.620] |
| | | (0.061) | | (0.082) | (0.095) | |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low GPA | 0.244*** | High GPA | 0.362*** | 0.118* | 0.090 [0.288] |
| | | (0.073) | | (0.062) | (0.067) | |
| Identity and self-concept | Low GPA | 0.101** | High GPA | 0.099*** | -0.003 | 0.965 [0.962] |
| | | (0.047) | | (0.031) | (0.062) | |
| Intentions and aspirations | Low GPA | -0.016 | High GPA | -0.054 | -0.038 | 0.608 [0.800] |
| | | (0.085) | | (0.063) | (0.073) | |
| Gender stereotypes | Low GPA | 0.099** | High GPA | -0.007 | -0.105 | 0.119 [0.288] |
| | | (0.047) | | (0.053) | (0.065) | |

*Notes:* The moderator is baseline grade point average; subgroup labels follow the preregistered cutoff shown in the table. Panel A reports economics applications, Panel B follow-up survey outcomes, and Panel C immediate post-intervention survey outcomes. Each row is estimated from a pooled model containing treated-school assignment, the displayed subgroup indicator, and their interaction. Effect low and Effect high are subgroup-specific linear combinations; 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 expressed in baseline control-group standard deviations and oriented so that positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100; effects, standard errors, and control means are therefore reported in percentage points. The conventional $p$-value tests the interaction. The bracketed wild $q$-value applies a false-discovery-rate adjustment to null-imposed school wild-bootstrap $p$-values within the displayed moderator family. Standard errors 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 S32: Heterogeneity by Baseline Classroom Female Share

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low female share | 1.769** | High female share | -0.023 | -1.792* | 0.072 [0.418] |
| | | (0.818) | | (0.421) | (0.958) | |
| Administrative applications | Low female share | 0.964 | High female share | 0.160 | -0.804 | 0.291 [0.762] |
| | | (0.617) | | (0.318) | (0.746) | |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low female share | 0.221** | High female share | 0.130* | -0.091 | 0.377 [0.716] |
| | | (0.082) | | (0.066) | (0.101) | |
| Identity and self-concept | Low female share | -0.033 | High female share | 0.110** | 0.143* | 0.066 [0.445] |
| | | (0.060) | | (0.053) | (0.075) | |
| Intentions and aspirations | Low female share | 0.052 | High female share | -0.001 | -0.053 | 0.490 [0.724] |
| | | (0.071) | | (0.035) | (0.076) | |
| Gender stereotypes | Low female share | 0.140 | High female share | -0.030 | -0.170 | 0.103 [0.469] |
| | | (0.083) | | (0.070) | (0.101) | |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low female share | 0.423*** | High female share | 0.194** | -0.228** | 0.023 [0.187] |
| | | (0.076) | | (0.072) | (0.094) | |
| Identity and self-concept | Low female share | 0.157*** | High female share | 0.047 | -0.110* | 0.071 [0.344] |
| | | (0.040) | | (0.035) | (0.059) | |
| Intentions and aspirations | Low female share | 0.076 | High female share | -0.140 | -0.216** | 0.016 [0.187] |
| | | (0.059) | | (0.086) | (0.084) | |
| Gender stereotypes | Low female share | 0.013 | High female share | 0.082** | 0.069 | 0.237 [0.602] |
| | | (0.054) | | (0.037) | (0.057) | |

*Notes:* The moderator is baseline classroom female share; the moderator is constructed before assignment and subgroup labels follow the prespecified split shown in the table. Panel A reports economics applications, Panel B follow-up survey outcomes, and Panel C immediate post-intervention survey outcomes. Each row is estimated from a pooled model containing treated-school assignment, the displayed subgroup indicator, and their interaction. Effect low and Effect high are subgroup-specific linear combinations; 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 expressed in baseline control-group standard deviations and oriented so that positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100; effects, standard errors, and control means are therefore reported in percentage points. The conventional $p$-value tests the interaction. The bracketed wild $q$-value applies a false-discovery-rate adjustment to null-imposed school wild-bootstrap $p$-values within the displayed moderator family. Some assignment-by-subgroup cells contain few school clusters, so the interaction estimates should be interpreted cautiously. Standard errors 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 S33: Heterogeneity by Baseline Classroom Mean GPA

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Lower two classroom terciles | 0.440 (0.343) | Top classroom GPA tercile | 0.767 (1.014) | 0.327 (1.106) | 0.770 [0.903] |
| Administrative applications | Lower two classroom terciles | 0.368 (0.334) | Top classroom GPA tercile | 0.479 (0.697) | 0.111 (0.837) | 0.895 [0.903] |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Lower two classroom terciles | 0.156** (0.060) | Top classroom GPA tercile | 0.218*** (0.072) | 0.062 (0.081) | 0.451 [0.724] |
| Identity and self-concept | Lower two classroom terciles | 0.029 (0.047) | Top classroom GPA tercile | 0.034 (0.078) | 0.004 (0.074) | 0.952 [0.992] |
| Intentions and aspirations | Lower two classroom terciles | 0.016 (0.044) | Top classroom GPA tercile | 0.008 (0.088) | -0.008 (0.106) | 0.942 [0.992] |
| Gender stereotypes | Lower two classroom terciles | 0.046 (0.055) | Top classroom GPA tercile | 0.064 (0.115) | 0.018 (0.114) | 0.877 [0.992] |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Lower two classroom terciles | 0.287*** (0.071) | Top classroom GPA tercile | 0.346*** (0.078) | 0.059 (0.096) | 0.542 [0.784] |
| Identity and self-concept | Lower two classroom terciles | 0.102*** (0.036) | Top classroom GPA tercile | 0.085 (0.053) | -0.017 (0.069) | 0.806 [0.855] |
| Intentions and aspirations | Lower two classroom terciles | -0.031 (0.079) | Top classroom GPA tercile | -0.055 (0.091) | -0.024 (0.116) | 0.839 [0.855] |
| Gender stereotypes | Lower two classroom terciles | 0.044 (0.038) | Top classroom GPA tercile | 0.071 (0.069) | 0.026 (0.070) | 0.713 [0.855] |

*Notes:* The moderator is baseline classroom mean GPA; the moderator is constructed before assignment and subgroup labels follow the prespecified split shown in the table. Panel A reports economics applications, Panel B follow-up survey outcomes, and Panel C immediate post-intervention survey outcomes. Each row is estimated from a pooled model containing treated-school assignment, the displayed subgroup indicator, and their interaction. Effect low and Effect high are subgroup-specific linear combinations; 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 expressed in baseline control-group standard deviations and oriented so that positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100; effects, standard errors, and control means are therefore reported in percentage points. The conventional $p$-value tests the interaction. The bracketed wild $q$-value applies a false-discovery-rate adjustment to null-imposed school wild-bootstrap $p$-values within the displayed moderator family. Some assignment-by-subgroup cells contain few school clusters, so the interaction estimates should be interpreted cautiously. Standard errors 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 S34: Heterogeneity by Baseline Classroom Gender Stereotypes

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low stereotypes | 0.350 | High stereotypes | 0.782 | 0.432 | 0.679 [0.903] |
| | | (0.513) | | (0.767) | (1.031) | |
| Administrative applications | Low stereotypes | 0.303 | High stereotypes | 0.549 | 0.246 | 0.755 [0.903] |
| | | (0.388) | | (0.570) | (0.779) | |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low stereotypes | 0.221*** | High stereotypes | 0.086 | -0.135 | 0.185 [0.545] |
| | | (0.078) | | (0.071) | (0.099) | |
| Identity and self-concept | Low stereotypes | 0.054 | High stereotypes | 0.021 | -0.033 | 0.591 [0.818] |
| | | (0.058) | | (0.054) | (0.060) | |
| Intentions and aspirations | Low stereotypes | -0.061 | High stereotypes | 0.053 | 0.114 | 0.166 [0.545] |
| | | (0.051) | | (0.056) | (0.080) | |
| Gender stereotypes | Low stereotypes | -0.099 | High stereotypes | 0.140* | 0.239** | 0.023 [0.336] |
| | | (0.078) | | (0.078) | (0.099) | |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low stereotypes | 0.367*** | High stereotypes | 0.203** | -0.164 | 0.104 [0.362] |
| | | (0.072) | | (0.085) | (0.098) | |
| Identity and self-concept | Low stereotypes | 0.152*** | High stereotypes | 0.031 | -0.120 | 0.124 [0.362] |
| | | (0.046) | | (0.048) | (0.076) | |
| Intentions and aspirations | Low stereotypes | 0.000 | High stereotypes | -0.091 | -0.091 | 0.395 [0.784] |
| | | (0.067) | | (0.101) | (0.105) | |
| Gender stereotypes | Low stereotypes | 0.016 | High stereotypes | 0.078 | 0.062 | 0.432 [0.784] |
| | | (0.054) | | (0.051) | (0.078) | |

*Notes:* The moderator is baseline classroom endorsement of the stereotype that men are naturally better at mathematics; subgroup labels follow the prespecified split shown in the table. Panel A reports economics applications, Panel B follow-up survey outcomes, and Panel C immediate post-intervention survey outcomes. Each row is estimated from a pooled model containing treated-school assignment, the displayed subgroup indicator, and their interaction. Effect low and Effect high are subgroup-specific linear combinations; 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 expressed in baseline control-group standard deviations and oriented so that positive values indicate more economics-oriented responses or less stereotyped views. Application outcomes are binary indicators multiplied by 100; effects, standard errors, and control means are therefore reported in percentage points. The conventional $p$-value tests the interaction. The bracketed wild $q$-value applies a false-discovery-rate adjustment to null-imposed school wild-bootstrap $p$-values within the displayed moderator family. Some assignment-by-subgroup cells contain few school clusters, so the interaction estimates should be interpreted cautiously. Standard errors 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 S35: Exploratory Heterogeneity by Classroom Teacher Gender

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Male teacher | 1.646 | Female teacher | 0.011 | -1.635 | 0.162 [0.569] |
| | | (0.975) | | (0.368) | (1.138) | |
| Administrative applications | Male teacher | 2.083** | Female teacher | -0.174 | -2.257** | 0.020 [0.418] |
| | | (0.887) | | (0.230) | (0.914) | |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Male teacher | 0.192** | Female teacher | 0.193*** | 0.001 | 0.992 [0.992] |
| | | (0.092) | | (0.057) | (0.086) | |
| Identity and self-concept | Male teacher | 0.167** | Female teacher | 0.027 | -0.140* | 0.060 [0.445] |
| | | (0.064) | | (0.048) | (0.071) | |
| Intentions and aspirations | Male teacher | -0.056 | Female teacher | 0.042 | 0.098 | 0.306 [0.716] |
| | | (0.064) | | (0.053) | (0.094) | |
| Gender stereotypes | Male teacher | -0.036 | Female teacher | 0.108 | 0.144 | 0.272 [0.716] |
| | | (0.098) | | (0.074) | (0.128) | |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Male teacher | 0.093 | Female teacher | 0.438*** | 0.345*** | $<$ 0.001 [0.112] |
| | | (0.065) | | (0.060) | (0.092) | |
| Identity and self-concept | Male teacher | 0.106** | Female teacher | 0.147*** | 0.041 | 0.468 [0.784] |
| | | (0.048) | | (0.028) | (0.056) | |
| Intentions and aspirations | Male teacher | -0.050 | Female teacher | 0.013 | 0.063 | 0.547 [0.784] |
| | | (0.107) | | (0.052) | (0.103) | |
| Gender stereotypes | Male teacher | 0.009 | Female teacher | 0.037 | 0.028 | 0.687 [0.855] |
| | | (0.062) | | (0.044) | (0.069) | |

*Notes:* Panel A reports administrative and gateway economics-application effects for women; Panels B–C report follow-up and immediate survey-index effects for both genders. Administrative applications cover Economics at the Universidad Autónoma de Nuevo León (UANL) and Licenciatura en Economía (LEC) at Tecnológico de Monterrey (Tec); broad applications additionally include Tec's Ciencias Sociales y Gobierno (ECSG) gateway. Application effects are in percentage points. Survey effects are in baseline control-group standard deviations, with higher values indicating more economics-oriented responses or less stereotyped views. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Each row comes from a pooled regression with teacher gender, school assignment, their interaction, and randomization-block fixed effects. Survey models also include the corresponding baseline index. Low denotes a male classroom teacher, High a female teacher, and High - low is the interaction coefficient. Standard errors are clustered by school. The $p$-value tests this difference; bracketed wild-bootstrap $q$-values retain the Benjamini–Hochberg adjustment across all classroom-context moderators and outcomes within each wave and outcome type, not only the teacher-gender rows displayed here. The sample consists of final-year baseline respondents with observed teacher gender and regression variables, with the additional restriction to women for applications. This exploratory extension compares intervention effects across teaching contexts and was not registered in the pre-analysis plan (PAP). Teacher gender was not randomized; the interaction does not identify the effect of assigning a female teacher or isolate a role-model mechanism. Sources are teacher records, student surveys, and institutional application records. Standard errors 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 S36: Gendered Heterogeneity by Baseline Beliefs and Interest

| Wave | Outcome family | Men, low | Women, low | Men, high | Women, high | Difference in gender effects | $p$-value [$q$-value] |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A. Moderator: Baseline beliefs about economics* | | | | | | | |
| Follow-up | Beliefs about economics | 0.234* | 0.117 | 0.044 | 0.282*** | 0.356** | 0.024 [0.097] |
| | | (0.127) | (0.073) | (0.073) | (0.057) | (0.149) | |
| Follow-up | Identity and self-concept | 0.026 | 0.108 | -0.070 | 0.071 | 0.059 | 0.673 [0.673] |
| | | (0.070) | (0.078) | (0.062) | (0.072) | (0.137) | |
| Follow-up | Intentions and aspirations | 0.035 | -0.020 | 0.029 | 0.072 | 0.099 | 0.509 [0.673] |
| | | (0.088) | (0.085) | (0.110) | (0.065) | (0.147) | |
| Follow-up | Gender stereotypes | 0.148 | -0.004 | 0.089 | 0.042 | 0.106 | 0.527 [0.673] |
| | | (0.138) | (0.081) | (0.085) | (0.068) | (0.165) | |
| Immediate | Beliefs about economics | 0.358*** | 0.417*** | 0.195** | 0.261*** | 0.007 | 0.945 [0.945] |
| | | (0.098) | (0.075) | (0.077) | (0.082) | (0.101) | |
| Immediate | Identity and self-concept | -0.054 | 0.182*** | 0.038 | 0.124** | -0.149 | 0.171 [0.618] |
| | | (0.061) | (0.050) | (0.064) | (0.051) | (0.106) | |
| Immediate | Intentions and aspirations | -0.097 | -0.038 | 0.032 | -0.006 | -0.097 | 0.386 [0.618] |
| | | (0.097) | (0.068) | (0.101) | (0.060) | (0.111) | |
| Immediate | Gender stereotypes | 0.136 | 0.003 | 0.079 | 0.033 | 0.087 | 0.464 [0.618] |
| | | (0.087) | (0.066) | (0.069) | (0.056) | (0.117) | |

*Continued on next page*

Table S36 continued

| Wave | Outcome family | Men, low | Women, low | Men, high | Women, high | Difference in gender effects | $p$-value [$q$-value] |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B. Moderator: Baseline interest in economics* | | | | | | | |
| Follow-up | Beliefs about economics | 0.087 | 0.207*** | 0.145* | 0.173*** | -0.092 | 0.666 [0.904] |
| | | (0.166) | (0.073) | (0.077) | (0.061) | (0.211) | |
| Follow-up | Identity and self-concept | 0.082 | 0.146 | -0.058 | 0.025 | 0.019 | 0.904 [0.904] |
| | | (0.106) | (0.093) | (0.068) | (0.056) | (0.155) | |
| Follow-up | Intentions and aspirations | 0.142 | -0.021 | 0.005 | 0.060 | 0.218 | 0.289 [0.904] |
| | | (0.161) | (0.054) | (0.072) | (0.082) | (0.202) | |
| Follow-up | Gender stereotypes | 0.038 | -0.017 | 0.135 | 0.049 | -0.031 | 0.869 [0.904] |
| | | (0.114) | (0.100) | (0.094) | (0.083) | (0.184) | |
| Immediate | Beliefs about economics | 0.440*** | 0.458*** | 0.221*** | 0.234*** | -0.004 | 0.977 [0.977] |
| | | (0.137) | (0.089) | (0.071) | (0.082) | (0.148) | |
| Immediate | Identity and self-concept | -0.017 | 0.102** | 0.025 | 0.182*** | 0.037 | 0.739 [0.977] |
| | | (0.068) | (0.046) | (0.058) | (0.038) | (0.111) | |
| Immediate | Intentions and aspirations | -0.073 | -0.039 | -0.012 | -0.040 | -0.061 | 0.727 [0.977] |
| | | (0.088) | (0.089) | (0.104) | (0.064) | (0.174) | |
| Immediate | Gender stereotypes | 0.097 | 0.010 | 0.098 | 0.026 | 0.015 | 0.921 [0.977] |
| | | (0.100) | (0.064) | (0.071) | (0.059) | (0.147) | |

*Notes:* The outcomes are immediate and follow-up survey indices. Panel A splits students by their baseline economics-beliefs index and Panel B by baseline interest in economics. High denotes a value at or above the median among final-year baseline respondents; Low denotes a value below that median. Survey effects are in baseline control-group standard deviations, with higher values indicating more economics-oriented responses or less stereotyped views. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. For each row, a pooled regression includes assignment, gender, the binary moderator, all two-way interactions, their three-way interaction, randomization-block fixed effects, and the corresponding baseline index. Columns (2)–(5) report assignment effects for men and women in the low and high groups as linear combinations from this model. Column (6), Difference in gender effects, subtracts the women-minus-men treatment-effect difference in the low group from the corresponding difference in the high group. A positive value indicates a larger relative response of women in the high group. Standard errors are clustered by school. The $p$-value in column (7) tests this three-way interaction. Bracketed $q$-values apply Benjamini–Hochberg adjustment to conventional school-clustered $p$-values across the four outcome domains separately for each moderator and survey wave. They are not wild-bootstrap values or a joint adjustment across the 32 interactions for the four moderators and two waves. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. Observations with a missing moderator are excluded. These selected three-way interactions are exploratory, not causal effects of changing the moderator. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S37: Gendered Heterogeneity by Baseline Mathematics Self-Assessment and Teacher Gender

| Wave | Outcome family | Men, low | Women, low | Men, high | Women, high | Difference in gender effects | $p$-value [$q$-value] |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A. Moderator: Baseline mathematics self-assessment* | | | | | | | |
| Follow-up | Beliefs about economics | 0.101 | 0.181*** | 0.146* | 0.186** | -0.038 | 0.806 [0.806] |
| | | (0.130) | (0.047) | (0.082) | (0.074) | (0.155) | |
| Follow-up | Identity and self-concept | -0.099 | 0.069 | 0.015 | 0.093 | -0.089 | 0.493 [0.779] |
| | | (0.093) | (0.066) | (0.067) | (0.074) | (0.128) | |
| Follow-up | Intentions and aspirations | 0.066 | 0.069 | 0.019 | -0.066 | -0.087 | 0.584 [0.779] |
| | | (0.125) | (0.064) | (0.088) | (0.070) | (0.157) | |
| Follow-up | Gender stereotypes | 0.033 | 0.058 | 0.147 | -0.061 | -0.233 | 0.207 [0.779] |
| | | (0.164) | (0.098) | (0.093) | (0.075) | (0.180) | |
| Immediate | Beliefs about economics | 0.161 | 0.378*** | 0.327*** | 0.292*** | -0.253** | 0.041 [0.165] |
| | | (0.102) | (0.070) | (0.072) | (0.099) | (0.118) | |
| Immediate | Identity and self-concept | 0.042 | 0.147*** | 0.004 | 0.153*** | 0.045 | 0.661 [0.818] |
| | | (0.071) | (0.048) | (0.044) | (0.048) | (0.101) | |
| Immediate | Intentions and aspirations | -0.006 | -0.039 | -0.039 | -0.041 | 0.031 | 0.818 [0.818] |
| | | (0.126) | (0.082) | (0.076) | (0.067) | (0.133) | |
| Immediate | Gender stereotypes | 0.183* | -0.003 | 0.054 | 0.037 | 0.169 | 0.171 [0.342] |
| | | (0.098) | (0.062) | (0.063) | (0.062) | (0.120) | |

*Continued on next page*

Table S37 continued

| Wave | Outcome family | Men, low | Women, low | Men, high | Women, high | Difference in gender effects | $p$-value [$q$-value] |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B. Moderator: Classroom teacher gender* | | | | | | | |
| Follow-up | Beliefs about economics | 0.251* | 0.127* | 0.136* | 0.227*** | 0.215 | 0.220 [0.380] |
| | | (0.142) | (0.069) | (0.069) | (0.074) | (0.171) | |
| Follow-up | Identity and self-concept | 0.053 | 0.241*** | -0.016 | 0.074 | -0.098 | 0.255 [0.380] |
| | | (0.077) | (0.070) | (0.053) | (0.065) | (0.085) | |
| Follow-up | Intentions and aspirations | 0.026 | -0.113 | 0.040 | 0.057 | 0.156 | 0.285 [0.380] |
| | | (0.125) | (0.070) | (0.085) | (0.074) | (0.143) | |
| Follow-up | Gender stereotypes | 0.064 | -0.079 | 0.133 | 0.083 | 0.093 | 0.636 [0.636] |
| | | (0.161) | (0.118) | (0.109) | (0.084) | (0.194) | |
| Immediate | Beliefs about economics | 0.174* | 0.052 | 0.367*** | 0.505*** | 0.261 | 0.111 [0.243] |
| | | (0.088) | (0.087) | (0.076) | (0.071) | (0.159) | |
| Immediate | Identity and self-concept | 0.055 | 0.152** | 0.079* | 0.192*** | 0.016 | 0.890 [0.890] |
| | | (0.092) | (0.060) | (0.043) | (0.034) | (0.116) | |
| Immediate | Intentions and aspirations | -0.041 | -0.018 | 0.028 | 0.005 | -0.047 | 0.757 [0.890] |
| | | (0.088) | (0.134) | (0.079) | (0.057) | (0.150) | |
| Immediate | Gender stereotypes | -0.111 | 0.080 | 0.084 | -0.001 | -0.276 | 0.121 [0.243] |
| | | (0.131) | (0.070) | (0.054) | (0.057) | (0.173) | |

*Notes:* The outcomes are immediate and follow-up survey indices. In Panel A, High denotes baseline mathematics self-assessment at or above the median among final-year baseline respondents, and Low denotes values below it. In Panel B, High denotes a female classroom teacher and Low a male teacher; it is not a median split. Teacher gender was not randomized. Survey effects are in baseline control-group standard deviations, with higher values indicating more economics-oriented responses or less stereotyped views. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. For each row, a pooled regression includes assignment, gender, the binary moderator, all two-way interactions, their three-way interaction, randomization-block fixed effects, and the corresponding baseline index. Columns (2)–(5) report assignment effects for men and women in the low and high groups as linear combinations from this model. Column (6), Difference in gender effects, subtracts the women-minus-men treatment-effect difference in the low group from the corresponding difference in the high group. A positive value indicates a larger relative response of women in the high group. Standard errors are clustered by school. The $p$-value in column (7) tests this three-way interaction. Bracketed $q$-values apply Benjamini–Hochberg adjustment to conventional school-clustered $p$-values across the four outcome domains separately for each moderator and survey wave. They are not wild-bootstrap values or a joint adjustment across the 32 interactions for the four moderators and two waves. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. Observations with a missing moderator are excluded. These selected three-way interactions are exploratory, not causal effects of changing the moderator. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S38: Heterogeneity by Altruism

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low Altruism | 0.500 | High Altruism | 0.804 | 0.303 | 0.693 [0.967] |
| | | (0.518) | | (0.550) | (0.760) | |
| Administrative applications | Low Altruism | 0.328 | High Altruism | 0.378 | 0.050 | 0.936 [0.981] |
| | | (0.468) | | (0.352) | (0.613) | |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low Altruism | 0.132** | High Altruism | 0.206*** | 0.074 | 0.319 [0.810] |
| | | (0.059) | | (0.060) | (0.073) | |
| Identity and self-concept | Low Altruism | 0.041 | High Altruism | 0.027 | -0.014 | 0.828 [0.936] |
| | | (0.052) | | (0.060) | (0.064) | |
| Intentions and aspirations | Low Altruism | -0.008 | High Altruism | 0.028 | 0.036 | 0.796 [0.925] |
| | | (0.073) | | (0.085) | (0.138) | |
| Gender stereotypes | Low Altruism | 0.094 | High Altruism | 0.021 | -0.073 | 0.395 [0.810] |
| | | (0.066) | | (0.073) | (0.085) | |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low Altruism | 0.185*** | High Altruism | 0.409*** | 0.224*** | 0.002 [0.179] |
| | | (0.060) | | (0.067) | (0.067) | |
| Identity and self-concept | Low Altruism | 0.066 | High Altruism | 0.128*** | 0.062 | 0.321 [0.819] |
| | | (0.046) | | (0.031) | (0.061) | |
| Intentions and aspirations | Low Altruism | -0.070 | High Altruism | 0.005 | 0.075 | 0.507 [0.819] |
| | | (0.109) | | (0.034) | (0.112) | |
| Gender stereotypes | Low Altruism | 0.148*** | High Altruism | -0.027 | -0.175** | 0.015 [0.340] |
| | | (0.049) | | (0.052) | (0.068) | |

*Notes:* The grouping variable is altruism, reported in the immediate post-intervention survey. Low denotes values below the control-arm median; High includes values at or above it. Panel A uses women in the final semester and reports administrative and gateway economics applications. Panels B and C use all final-semester baseline respondents observed at follow-up and immediately after intervention, respectively. Each row comes from a pooled model containing treated-school assignment, subgroup membership, and their interaction. Effect low and Effect high are the assignment coefficients for the two groups; High - low is the interaction coefficient. All models include randomization-block fixed effects and school-clustered standard errors; survey models also control for the matching baseline index. Survey effects are in baseline control-group standard deviations and application effects in percentage points. The $p$-value uses school-clustered inference; the bracketed wild $q$-value retains the original adjustment of wild-bootstrap interaction tests across the fourteen grouping variables within wave and outcome-type families. Splitting the presentation does not redefine those families. These non-prespecified, post-assignment splits are descriptive, not causal effects within predetermined subgroups or identified mediation channels. Standard errors 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 S39: Heterogeneity by Quantitative Altruism

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low Quantitative altruism | 0.824 | High Quantitative altruism | 0.699* | -0.125 | 0.894 [0.981] |
| | | (0.878) | | (0.376) | (0.930) | |
| Administrative applications | Low Quantitative altruism | 0.393 | High Quantitative altruism | 0.394 | 0.001 | 0.999 [0.999] |
| | | (0.617) | | (0.281) | (0.666) | |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low Quantitative altruism | 0.275*** | High Quantitative altruism | 0.138** | -0.137* | 0.053 [0.761] |
| | | (0.057) | | (0.058) | (0.068) | |
| Identity and self-concept | Low Quantitative altruism | -0.002 | High Quantitative altruism | 0.044 | 0.046 | 0.461 [0.810] |
| | | (0.063) | | (0.053) | (0.061) | |
| Intentions and aspirations | Low Quantitative altruism | 0.071 | High Quantitative altruism | -0.032 | -0.104 | 0.110 [0.761] |
| | | (0.053) | | (0.049) | (0.063) | |
| Gender stereotypes | Low Quantitative altruism | 0.039 | High Quantitative altruism | 0.090 | 0.050 | 0.541 [0.820] |
| | | (0.062) | | (0.072) | (0.081) | |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low Quantitative altruism | 0.297*** | High Quantitative altruism | 0.329*** | 0.031 | 0.710 [0.886] |
| | | (0.084) | | (0.063) | (0.083) | |
| Identity and self-concept | Low Quantitative altruism | 0.076 | High Quantitative altruism | 0.121*** | 0.045 | 0.454 [0.819] |
| | | (0.046) | | (0.032) | (0.059) | |
| Intentions and aspirations | Low Quantitative altruism | -0.129 | High Quantitative altruism | 0.006 | 0.135 | 0.343 [0.819] |
| | | (0.137) | | (0.051) | (0.140) | |
| Gender stereotypes | Low Quantitative altruism | 0.123* | High Quantitative altruism | 0.019 | -0.103 | 0.262 [0.819] |
| | | (0.063) | | (0.056) | (0.090) | |

*Notes:* The grouping variable is quantitative altruism, reported in the immediate post-intervention survey. Low denotes values below the control-arm median; High includes values at or above it. Panel A uses women in the final semester and reports administrative and gateway economics applications. Panels B and C use all final-semester baseline respondents observed at follow-up and immediately after intervention, respectively. Each row comes from a pooled model containing treated-school assignment, subgroup membership, and their interaction. Effect low and Effect high are the assignment coefficients for the two groups; High - low is the interaction coefficient. All models include randomization-block fixed effects and school-clustered standard errors; survey models also control for the matching baseline index. Survey effects are in baseline control-group standard deviations and application effects in percentage points. The $p$-value uses school-clustered inference; the bracketed wild $q$-value retains the original adjustment of wild-bootstrap interaction tests across the fourteen grouping variables within wave and outcome-type families. Splitting the presentation does not redefine those families. These non-prespecified, post-assignment splits are descriptive, not causal effects within predetermined subgroups or identified mediation channels. Standard errors 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 S40: Heterogeneity by Generalized Trust

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low Generalized trust | 0.554 (0.437) | High Generalized trust | 0.816 (0.643) | 0.262 (0.779) | 0.739 [0.967] |
| Administrative applications | Low Generalized trust | 0.478 (0.409) | High Generalized trust | 0.236 (0.383) | -0.242 (0.586) | 0.683 [0.967] |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low Generalized trust | 0.172** (0.063) | High Generalized trust | 0.169*** (0.054) | -0.004 (0.068) | 0.957 [0.993] |
| Identity and self-concept | Low Generalized trust | 0.062 (0.068) | High Generalized trust | -0.003 (0.047) | -0.065 (0.072) | 0.370 [0.810] |
| Intentions and aspirations | Low Generalized trust | 0.009 (0.064) | High Generalized trust | 0.012 (0.063) | 0.003 (0.099) | 0.973 [0.993] |
| Gender stereotypes | Low Generalized trust | 0.107 (0.066) | High Generalized trust | 0.012 (0.073) | -0.094 (0.079) | 0.240 [0.810] |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low Generalized trust | 0.236*** (0.062) | High Generalized trust | 0.364*** (0.069) | 0.128* (0.063) | 0.053 [0.625] |
| Identity and self-concept | Low Generalized trust | 0.126*** (0.041) | High Generalized trust | 0.077** (0.036) | -0.049 (0.063) | 0.444 [0.819] |
| Intentions and aspirations | Low Generalized trust | -0.063 (0.111) | High Generalized trust | -0.008 (0.042) | 0.055 (0.113) | 0.629 [0.839] |
| Gender stereotypes | Low Generalized trust | 0.023 (0.038) | High Generalized trust | 0.074 (0.053) | 0.051 (0.059) | 0.394 [0.819] |

*Notes:* The grouping variable is generalized trust, reported in the immediate post-intervention survey. Low denotes values below the control-arm median; High includes values at or above it. Panel A uses women in the final semester and reports administrative and gateway economics applications. Panels B and C use all final-semester baseline respondents observed at follow-up and immediately after intervention, respectively. Each row comes from a pooled model containing treated-school assignment, subgroup membership, and their interaction. Effect low and Effect high are the assignment coefficients for the two groups; High - low is the interaction coefficient. All models include randomization-block fixed effects and school-clustered standard errors; survey models also control for the matching baseline index. Survey effects are in baseline control-group standard deviations and application effects in percentage points. The $p$-value uses school-clustered inference; the bracketed wild $q$-value retains the original adjustment of wild-bootstrap interaction tests across the fourteen grouping variables within wave and outcome-type families. Splitting the presentation does not redefine those families. These non-prespecified, post-assignment splits are descriptive, not causal effects within predetermined subgroups or identified mediation channels. Standard errors 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 S41: Heterogeneity by Willingness to Wait

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low Patience willingness | 0.908* (0.473) | High Patience willingness | 0.464 (0.670) | -0.445 (0.882) | 0.618 [0.967] |
| Administrative applications | Low Patience willingness | 0.278 (0.292) | High Patience willingness | 0.434 (0.431) | 0.155 (0.499) | 0.758 [0.967] |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low Patience willingness | 0.229*** (0.070) | High Patience willingness | 0.128** (0.048) | -0.101 (0.064) | 0.127 [0.761] |
| Identity and self-concept | Low Patience willingness | 0.008 (0.058) | High Patience willingness | 0.053 (0.059) | 0.045 (0.074) | 0.546 [0.820] |
| Intentions and aspirations | Low Patience willingness | 0.027 (0.049) | High Patience willingness | 0.001 (0.067) | -0.025 (0.089) | 0.776 [0.906] |
| Gender stereotypes | Low Patience willingness | 0.098 (0.064) | High Patience willingness | 0.026 (0.083) | -0.071 (0.097) | 0.469 [0.810] |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low Patience willingness | 0.291*** (0.057) | High Patience willingness | 0.308*** (0.068) | 0.016 (0.059) | 0.786 [0.946] |
| Identity and self-concept | Low Patience willingness | 0.130** (0.050) | High Patience willingness | 0.074** (0.032) | -0.056 (0.068) | 0.418 [0.819] |
| Intentions and aspirations | Low Patience willingness | -0.070 (0.121) | High Patience willingness | -0.005 (0.036) | 0.065 (0.128) | 0.615 [0.824] |
| Gender stereotypes | Low Patience willingness | 0.058 (0.050) | High Patience willingness | 0.046 (0.054) | -0.012 (0.075) | 0.874 [0.947] |

*Notes:* The grouping variable is willingness to wait, reported in the immediate post-intervention survey. Low denotes values below the control-arm median; High includes values at or above it. Panel A uses women in the final semester and reports administrative and gateway economics applications. Panels B and C use all final-semester baseline respondents observed at follow-up and immediately after intervention, respectively. Each row comes from a pooled model containing treated-school assignment, subgroup membership, and their interaction. Effect low and Effect high are the assignment coefficients for the two groups; High - low is the interaction coefficient. All models include randomization-block fixed effects and school-clustered standard errors; survey models also control for the matching baseline index. Survey effects are in baseline control-group standard deviations and application effects in percentage points. The $p$-value uses school-clustered inference; the bracketed wild $q$-value retains the original adjustment of wild-bootstrap interaction tests across the fourteen grouping variables within wave and outcome-type families. Splitting the presentation does not redefine those families. These non-prespecified, post-assignment splits are descriptive, not causal effects within predetermined subgroups or identified mediation channels. Standard errors 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 S42: Heterogeneity by Patience Valuation

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low Patience valuation task | 0.637 (0.615) | High Patience valuation task | 0.849 (0.536) | 0.212 (0.829) | 0.800 [0.976] |
| Administrative applications | Low Patience valuation task | 0.305 (0.287) | High Patience valuation task | 0.512 (0.466) | 0.207 (0.544) | 0.707 [0.967] |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low Patience valuation task | 0.201*** (0.063) | High Patience valuation task | 0.201*** (0.061) | 0.000 (0.084) | 0.999 [0.999] |
| Identity and self-concept | Low Patience valuation task | -0.020 (0.052) | High Patience valuation task | 0.067 (0.054) | 0.087* (0.046) | 0.067 [0.761] |
| Intentions and aspirations | Low Patience valuation task | 0.021 (0.054) | High Patience valuation task | 0.005 (0.072) | -0.016 (0.101) | 0.875 [0.947] |
| Gender stereotypes | Low Patience valuation task | 0.105 (0.076) | High Patience valuation task | 0.049 (0.063) | -0.056 (0.085) | 0.517 [0.810] |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low Patience valuation task | 0.274*** (0.069) | High Patience valuation task | 0.365*** (0.069) | 0.090 (0.074) | 0.233 [0.819] |
| Identity and self-concept | Low Patience valuation task | 0.043 (0.054) | High Patience valuation task | 0.167*** (0.032) | 0.123 (0.074) | 0.108 [0.819] |
| Intentions and aspirations | Low Patience valuation task | -0.094 (0.117) | High Patience valuation task | 0.008 (0.043) | 0.102 (0.113) | 0.375 [0.819] |
| Gender stereotypes | Low Patience valuation task | 0.087 (0.059) | High Patience valuation task | 0.031 (0.055) | -0.056 (0.085) | 0.516 [0.819] |

*Notes:* The grouping variable is the patience valuation task, reported in the immediate post-intervention survey. Low denotes values below the control-arm median; High includes values at or above it. Panel A uses women in the final semester and reports administrative and gateway economics applications. Panels B and C use all final-semester baseline respondents observed at follow-up and immediately after intervention, respectively. Each row comes from a pooled model containing treated-school assignment, subgroup membership, and their interaction. Effect low and Effect high are the assignment coefficients for the two groups; High - low is the interaction coefficient. All models include randomization-block fixed effects and school-clustered standard errors; survey models also control for the matching baseline index. Survey effects are in baseline control-group standard deviations and application effects in percentage points. The $p$-value uses school-clustered inference; the bracketed wild $q$-value retains the original adjustment of wild-bootstrap interaction tests across the fourteen grouping variables within wave and outcome-type families. Splitting the presentation does not redefine those families. These non-prespecified, post-assignment splits are descriptive, not causal effects within predetermined subgroups or identified mediation channels. Standard errors 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 S43: Heterogeneity by Procrastination

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low Procrastination | 1.022** (0.438) | High Procrastination | 0.359 (0.487) | -0.664 (0.534) | 0.225 [0.652] |
| Administrative applications | Low Procrastination | 0.708* (0.388) | High Procrastination | 0.027 (0.362) | -0.681 (0.527) | 0.207 [0.652] |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low Procrastination | 0.110** (0.046) | High Procrastination | 0.234*** (0.075) | 0.125 (0.083) | 0.146 [0.761] |
| Identity and self-concept | Low Procrastination | 0.006 (0.062) | High Procrastination | 0.057 (0.051) | 0.051 (0.066) | 0.447 [0.810] |
| Intentions and aspirations | Low Procrastination | -0.009 (0.048) | High Procrastination | 0.032 (0.067) | 0.041 (0.087) | 0.641 [0.858] |
| Gender stereotypes | Low Procrastination | 0.090 (0.067) | High Procrastination | 0.033 (0.069) | -0.057 (0.074) | 0.448 [0.810] |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low Procrastination | 0.295*** (0.085) | High Procrastination | 0.303*** (0.062) | 0.008 (0.088) | 0.928 [0.953] |
| Identity and self-concept | Low Procrastination | 0.095** (0.043) | High Procrastination | 0.100** (0.038) | 0.005 (0.063) | 0.934 [0.953] |
| Intentions and aspirations | Low Procrastination | -0.098 (0.112) | High Procrastination | 0.015 (0.047) | 0.113 (0.114) | 0.333 [0.819] |
| Gender stereotypes | Low Procrastination | 0.018 (0.034) | High Procrastination | 0.083 (0.062) | 0.064 (0.068) | 0.355 [0.819] |

*Notes:* The grouping variable is procrastination, reported in the immediate post-intervention survey. Low denotes values below the control-arm median; High includes values at or above it. Panel A uses women in the final semester and reports administrative and gateway economics applications. Panels B and C use all final-semester baseline respondents observed at follow-up and immediately after intervention, respectively. Each row comes from a pooled model containing treated-school assignment, subgroup membership, and their interaction. Effect low and Effect high are the assignment coefficients for the two groups; High - low is the interaction coefficient. All models include randomization-block fixed effects and school-clustered standard errors; survey models also control for the matching baseline index. Survey effects are in baseline control-group standard deviations and application effects in percentage points. The $p$-value uses school-clustered inference; the bracketed wild $q$-value retains the original adjustment of wild-bootstrap interaction tests across the fourteen grouping variables within wave and outcome-type families. Splitting the presentation does not redefine those families. These non-prespecified, post-assignment splits are descriptive, not causal effects within predetermined subgroups or identified mediation channels. Procrastination is reported as a time-related behavioral measure. Standard errors 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 S44: Heterogeneity by Willingness to Take Risks

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low Risk willingness | 1.222** (0.526) | High Risk willingness | 0.169 (0.551) | -1.052 (0.777) | 0.187 [0.652] |
| Administrative applications | Low Risk willingness | 0.716* (0.397) | High Risk willingness | 0.019 (0.389) | -0.697 (0.579) | 0.239 [0.652] |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low Risk willingness | 0.217*** (0.044) | High Risk willingness | 0.138* (0.068) | -0.079 (0.067) | 0.249 [0.810] |
| Identity and self-concept | Low Risk willingness | -0.014 (0.060) | High Risk willingness | 0.074 (0.053) | 0.088 (0.067) | 0.199 [0.810] |
| Intentions and aspirations | Low Risk willingness | -0.015 (0.049) | High Risk willingness | 0.035 (0.048) | 0.051 (0.055) | 0.370 [0.810] |
| Gender stereotypes | Low Risk willingness | 0.031 (0.063) | High Risk willingness | 0.079 (0.068) | 0.048 (0.066) | 0.470 [0.810] |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low Risk willingness | 0.215*** (0.074) | High Risk willingness | 0.373*** (0.056) | 0.158** (0.065) | 0.023 [0.428] |
| Identity and self-concept | Low Risk willingness | 0.135** (0.050) | High Risk willingness | 0.073* (0.036) | -0.062 (0.074) | 0.407 [0.819] |
| Intentions and aspirations | Low Risk willingness | -0.108 (0.117) | High Risk willingness | 0.031 (0.039) | 0.140 (0.124) | 0.269 [0.819] |
| Gender stereotypes | Low Risk willingness | 0.081 (0.051) | High Risk willingness | 0.027 (0.065) | -0.054 (0.093) | 0.565 [0.819] |

*Notes:* The grouping variable is willingness to take risks, reported in the immediate post-intervention survey. Low denotes values below the control-arm median; High includes values at or above it. Panel A uses women in the final semester and reports administrative and gateway economics applications. Panels B and C use all final-semester baseline respondents observed at follow-up and immediately after intervention, respectively. Each row comes from a pooled model containing treated-school assignment, subgroup membership, and their interaction. Effect low and Effect high are the assignment coefficients for the two groups; High - low is the interaction coefficient. All models include randomization-block fixed effects and school-clustered standard errors; survey models also control for the matching baseline index. Survey effects are in baseline control-group standard deviations and application effects in percentage points. The $p$-value uses school-clustered inference; the bracketed wild $q$-value retains the original adjustment of wild-bootstrap interaction tests across the fourteen grouping variables within wave and outcome-type families. Splitting the presentation does not redefine those families. These non-prespecified, post-assignment splits are descriptive, not causal effects within predetermined subgroups or identified mediation channels. Standard errors 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 S45: Heterogeneity by Risk Valuation

| Outcome | Low subgroup (1) | Effect low (2) | High subgroup (3) | Effect high (4) | High - low (5) | $p$-value [wild $q$-value] (6) |
|---|---|---|---|---|---|---|
| *Panel A. Economic applications* | | | | | | |
| Economics application: broad gateway | Low Risk valuation task | 0.803 (0.491) | High Risk valuation task | 0.667 (0.668) | -0.136 (0.841) | 0.873 [0.981] |
| Administrative applications | Low Risk valuation task | 0.284 (0.352) | High Risk valuation task | 0.516 (0.489) | 0.232 (0.621) | 0.711 [0.967] |
| *Panel B. Follow-up outcomes* | | | | | | |
| Beliefs about economics | Low Risk valuation task | 0.314*** (0.060) | High Risk valuation task | 0.088 (0.061) | -0.226*** (0.076) | 0.006 [0.207] |
| Identity and self-concept | Low Risk valuation task | 0.034 (0.053) | High Risk valuation task | 0.021 (0.053) | -0.013 (0.045) | 0.775 [0.906] |
| Intentions and aspirations | Low Risk valuation task | 0.042 (0.061) | High Risk valuation task | -0.010 (0.052) | -0.052 (0.080) | 0.518 [0.810] |
| Gender stereotypes | Low Risk valuation task | 0.089 (0.060) | High Risk valuation task | 0.052 (0.087) | -0.037 (0.102) | 0.717 [0.906] |
| *Panel C. Immediate post-intervention outcomes* | | | | | | |
| Beliefs about economics | Low Risk valuation task | 0.280*** (0.079) | High Risk valuation task | 0.353*** (0.062) | 0.073 (0.078) | 0.358 [0.819] |
| Identity and self-concept | Low Risk valuation task | 0.081 (0.054) | High Risk valuation task | 0.123*** (0.033) | 0.042 (0.077) | 0.591 [0.819] |
| Intentions and aspirations | Low Risk valuation task | -0.094 (0.125) | High Risk valuation task | -0.014 (0.033) | 0.081 (0.124) | 0.523 [0.819] |
| Gender stereotypes | Low Risk valuation task | 0.041 (0.049) | High Risk valuation task | 0.072 (0.048) | 0.031 (0.059) | 0.602 [0.819] |

*Notes:* The grouping variable is the risk valuation task, reported in the immediate post-intervention survey. Low denotes values below the control-arm median; High includes values at or above it. Panel A uses women in the final semester and reports administrative and gateway economics applications. Panels B and C use all final-semester baseline respondents observed at follow-up and immediately after intervention, respectively. Each row comes from a pooled model containing treated-school assignment, subgroup membership, and their interaction. Effect low and Effect high are the assignment coefficients for the two groups; High - low is the interaction coefficient. All models include randomization-block fixed effects and school-clustered standard errors; survey models also control for the matching baseline index. Survey effects are in baseline control-group standard deviations and application effects in percentage points. The $p$-value uses school-clustered inference; the bracketed wild $q$-value retains the original adjustment of wild-bootstrap interaction tests across the fourteen grouping variables within wave and outcome-type families. Splitting the presentation does not redefine those families. These non-prespecified, post-assignment splits are descriptive, not causal effects within predetermined subgroups or identified mediation channels. Standard errors 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 S46: Machine-Learning Heterogeneity in Follow-up Survey Outcomes

| Model component | Estimate (1) | $p$-value (2) |
|---|---|---|
| *Panel A. Intentions and aspirations* | | |
| Orthogonal-score ATE | 0.029 | 0.522 |
| | (0.045) | |
| LASSO CATE calibration slope | -4.876* | 0.065 |
| | (2.532) | |
| LASSO top-minus-bottom CATE group | -0.212 | 0.163 |
| | (0.147) | |
| Random-forest CATE calibration slope | -1.218 | 0.154 |
| | (0.830) | |
| Random-forest top-minus-bottom CATE group | -0.259 | 0.188 |
| | (0.192) | |
| *Panel B. Beliefs about economics* | | |
| Orthogonal-score ATE | 0.187*** | 0.008 |
| | (0.065) | |
| LASSO CATE calibration slope | -0.174 | 0.486 |
| | (0.247) | |
| LASSO top-minus-bottom CATE group | -0.274 | 0.193 |
| | (0.205) | |
| Random-forest CATE calibration slope | -1.808** | 0.033 |
| | (0.805) | |
| Random-forest top-minus-bottom CATE group | -0.114 | 0.473 |
| | (0.157) | |
| *Panel C. Identity and self-concept* | | |
| Orthogonal-score ATE | 0.049 | 0.266 |
| | (0.043) | |
| LASSO CATE calibration slope | -4.244** | 0.011 |
| | (1.544) | |
| LASSO top-minus-bottom CATE group | -0.316** | 0.045 |
| | (0.151) | |
| Random-forest CATE calibration slope | 0.404 | 0.649 |
| | (0.878) | |
| Random-forest top-minus-bottom CATE group | 0.102 | 0.632 |
| | (0.211) | |
| *Panel D. Gender stereotypes* | | |
| Orthogonal-score ATE | 0.066 | 0.349 |
| | (0.069) | |
| LASSO CATE calibration slope | -4.308*** | 0.002 |
| | (1.279) | |
| LASSO top-minus-bottom CATE group | -0.391*** | 0.004 |
| | (0.123) | |
| Random-forest CATE calibration slope | 1.127* | 0.054 |
| | (0.558) | |
| Random-forest top-minus-bottom CATE group | 0.124 | 0.343 |
| | (0.128) | |

*Notes:* The LASSO uses school-level folds, while the native random-forest ranking uses observation-level folds. Panels correspond to the four follow-up survey-index families. The orthogonal-score ATE is the cross-fitted average treatment effect. The CATE calibration slope assesses whether realized orthogonal scores rise with the learner's predicted heterogeneity, and the top-minus-bottom CATE-group contrast compares observations assigned out of fold to the highest and lowest predicted-effect groups. Treatment propensities are fixed by the original blocked randomization, nuisance functions and CATE scores are estimated out of fold, and inference is clustered by school. Outcomes are in baseline control-group standard deviations. Decile rows and method-audit diagnostics remain in the machine-readable files. These exploratory tests assess predictive heterogeneity and do not identify causal effects of the characteristics used by a learner. The calibration $p$-value tests a zero slope; a slope of one indicates ideal calibration, while a negative slope indicates that realized orthogonal scores decline as predicted gains rise. The group-contrast $p$-value tests equality between the highest and lowest predicted-effect groups. These diagnostics assess the predictive ordering used in the subsequent gender summaries and characteristic profiles. Standard errors 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 S47: Machine-Learning Heterogeneity in Economics Applications

| Model component | Estimate (1) | $p$-value (2) |
|---|---|---|
| Orthogonal-score ATE | 0.535* | 0.069 |
| | (0.283) | |
| LASSO CATE calibration slope | -3.908*** | 0.004 |
| | (1.231) | |
| LASSO top-minus-bottom CATE group | -2.404*** | 0.003 |
| | (0.738) | |
| Random-forest CATE calibration slope | 0.608 | 0.607 |
| | (1.170) | |
| Random-forest top-minus-bottom CATE group | 3.879 | 0.222 |
| | (3.104) | |

*Notes:* The LASSO uses school-level folds, while the native random-forest ranking uses observation-level folds. The outcome is the administrative economics-application indicator for women in the final semester, expressed in percentage points. The orthogonal-score ATE is the cross-fitted average treatment effect. The CATE calibration slope assesses whether realized orthogonal scores rise with the learner's predicted heterogeneity, and the top-minus-bottom CATE-group contrast compares observations assigned out of fold to the highest and lowest predicted-effect groups. Treatment propensities are fixed by the blocked randomization, nuisance functions and CATE scores are estimated out of fold, and inference is clustered by school. Decile rows and method-audit diagnostics remain in the machine-readable files. Because the administrative outcome is rare, model calibration and subgroup contrasts are fragile and should be interpreted as exploratory. The calibration $p$-value tests a zero slope; a slope of one indicates ideal calibration, while a negative slope indicates that realized orthogonal scores decline as predicted gains rise. The group-contrast $p$-value tests equality between the highest and lowest predicted-effect groups. These diagnostics assess the predictive ordering used in the subsequent gender summaries and characteristic profiles. Standard errors 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 S48: Machine-Learning Heterogeneity by Student Gender

| CATE learner | Mean model-implied CATE (1) | $p$-value, women–men difference (2) |
|---|---|---|
| *Panel A. Women* | | |
| *Intentions and aspirations* | | |
| Cross-fitted LASSO | 0.027 | 0.932 |
| Generalized random forest | 0.000 | < 0.001 |
| *Beliefs about economics* | | |
| Cross-fitted LASSO | 0.183 | 0.209 |
| Generalized random forest | 0.154 | 0.124 |
| *Identity and self-concept* | | |
| Cross-fitted LASSO | 0.045 | 0.105 |
| Generalized random forest | 0.045 | < 0.001 |
| *Gender stereotypes* | | |
| Cross-fitted LASSO | 0.069 | 0.366 |
| Generalized random forest | 0.032 | < 0.001 |
| *Panel B. Men* | | |
| *Intentions and aspirations* | | |
| Cross-fitted LASSO | 0.027 | 0.932 |
| Generalized random forest | 0.020 | < 0.001 |
| *Beliefs about economics* | | |
| Cross-fitted LASSO | 0.173 | 0.209 |
| Generalized random forest | 0.159 | 0.124 |
| *Identity and self-concept* | | |
| Cross-fitted LASSO | 0.042 | 0.105 |
| Generalized random forest | 0.016 | < 0.001 |
| *Gender stereotypes* | | |
| Cross-fitted LASSO | 0.066 | 0.366 |
| Generalized random forest | 0.065 | < 0.001 |

*Notes:* The LASSO uses school-level folds, while the native random-forest ranking uses observation-level folds. Entries are subgroup means of out-of-fold, model-implied conditional average treatment-effect (CATE) scores for each follow-up survey family, in baseline control-group standard deviations. The women–men $p$-value tests equality of mean predicted scores by regressing each score on an indicator for women with school-clustered standard errors. Scores are estimated separately by the listed learner. These tests concern the predictions of the fitted models; they do not replace the randomized pooled treatment-by-gender interaction estimates in the main analysis. The calibration and group-effect tests in Table S46 do not consistently validate the predicted ordering. Accordingly, differences in mean scores describe how each learner predicts gains across genders and do not establish reliable differences in realized treatment effects. Generalized-random-forest rows appear where usable scores were produced. Significance is assessed using two-sided, unrounded $p$-values below 0.10.

Table S49: Predetermined Characteristics Associated with Model-Implied CATE Scores: Beliefs about Economics

| Predetermined characteristic | CATE-score gradient (1) | $p$-value (2) |
|---|---|---|
| *Panel A. Cross-fitted LASSO* | | |
| Household asset index | 0.009 | 0.179 |
| | (0.007) | |
| Works for pay | -0.008 | 0.335 |
| | (0.009) | |
| Female student | 0.005 | 0.208 |
| | (0.004) | |
| Mother completed university | 0.004 | 0.330 |
| | (0.004) | |
| *Panel B. Generalized random forest* | | |
| Classroom female share | -0.020*** | $<$ 0.001 |
| | (0.002) | |
| Family support for economics | 0.007*** | $<$ 0.001 |
| | (0.002) | |
| Self-assessed economics ability | -0.006*** | 0.003 |
| | (0.002) | |
| Baseline intentions index | 0.005** | 0.030 |
| | (0.002) | |

*Notes:* The LASSO uses school-level folds, while the native random-forest ranking uses observation-level folds. For the beliefs-about-economics outcome family, entries describe associations between predetermined characteristics and out-of-fold, model-implied conditional average treatment-effect (CATE) scores. Each row regresses the predicted score on one standardized characteristic, with school-clustered standard errors. A positive gradient means that a one-standard-deviation increase in the characteristic is associated with a larger predicted effect. Within each learner, the four largest absolute gradients among substantively interpretable features are shown; block indicators and other gradients remain in the machine-readable output. The $p$-values test associations with predicted scores, not randomized treatment-by-characteristic interactions. The profiles describe how the fitted learners sort students and should be read alongside the calibration and group-effect tests in Table S46. Those diagnostics do not consistently validate the predicted ordering, so these associations alone do not establish which students benefit most or identify causal mechanisms. Standard errors 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 S50: Predetermined Characteristics Associated with Model-Implied CATE Scores: Identity and Self-Concept

| Predetermined characteristic | CATE-score gradient (1) | $p$-value (2) |
|---|---|---|
| *Panel A. Cross-fitted LASSO* | | |
| Age | 0.003 | 0.339 |
| | (0.003) | |
| Classroom female share | 0.002 | 0.395 |
| | (0.003) | |
| Interest in history | -0.002 | 0.177 |
| | (0.001) | |
| Female student | 0.002 | 0.104 |
| | (0.001) | |
| *Panel B. Generalized random forest* | | |
| Baseline identity index | -0.038*** | $<$ 0.001 |
| | (0.003) | |
| Self-assessed economics ability | -0.037*** | $<$ 0.001 |
| | (0.003) | |
| Interest in economics | -0.037*** | $<$ 0.001 |
| | (0.003) | |
| Self-assessed mathematics ability | -0.028*** | $<$ 0.001 |
| | (0.004) | |

*Notes:* The LASSO uses school-level folds, while the native random-forest ranking uses observation-level folds. For the identity-and-self-concept outcome family, entries describe associations between predetermined characteristics and out-of-fold, model-implied conditional average treatment-effect (CATE) scores. Each row regresses the predicted score on one standardized characteristic, with school-clustered standard errors. A positive gradient means that a one-standard-deviation increase in the characteristic is associated with a larger predicted effect. Within each learner, the four largest absolute gradients among substantively interpretable features are shown; block indicators and other gradients remain in the machine-readable output. The $p$-values test associations with predicted scores, not randomized treatment-by-characteristic interactions. The profiles describe how the fitted learners sort students and should be read alongside the calibration and group-effect tests in Table S46. Those diagnostics do not consistently validate the predicted ordering, so these associations alone do not establish which students benefit most or identify causal mechanisms. Standard errors 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 S51: Predetermined Characteristics Associated with Model-Implied CATE Scores: Intentions and Aspirations

| Predetermined characteristic | CATE-score gradient (1) | $p$-value (2) |
|---|---|---|
| *Panel A. Cross-fitted LASSO* | | |
| Age | 0.002 | 0.267 |
| | (0.002) | |
| Classroom female share | 0.002 | 0.318 |
| | (0.002) | |
| Mother completed university | -0.001 | 0.298 |
| | (0.001) | |
| Father completed university | -0.001 | 0.343 |
| | (0.001) | |
| *Panel B. Generalized random forest* | | |
| Family support for economics | -0.034*** | $<$ 0.001 |
| | (0.003) | |
| Baseline intentions index | -0.027*** | $<$ 0.001 |
| | (0.002) | |
| Interest in history | 0.017*** | $<$ 0.001 |
| | (0.001) | |
| Interest in mathematics | 0.017*** | $<$ 0.001 |
| | (0.002) | |

*Notes:* The LASSO uses school-level folds, while the native random-forest ranking uses observation-level folds. For the intentions-and-aspirations outcome family, entries describe associations between predetermined characteristics and out-of-fold, model-implied conditional average treatment-effect (CATE) scores. Each row regresses the predicted score on one standardized characteristic, with school-clustered standard errors. A positive gradient means that a one-standard-deviation increase in the characteristic is associated with a larger predicted effect. Within each learner, the four largest absolute gradients among substantively interpretable features are shown; block indicators and other gradients remain in the machine-readable output. The $p$-values test associations with predicted scores, not randomized treatment-by-characteristic interactions. The profiles describe how the fitted learners sort students and should be read alongside the calibration and group-effect tests in Table S46. Those diagnostics do not consistently validate the predicted ordering, so these associations alone do not establish which students benefit most or identify causal mechanisms. Standard errors 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 S52: Predetermined Characteristics Associated with Model-Implied CATE Scores: Gender Stereotypes

| Predetermined characteristic | CATE-score gradient (1) | $p$-value (2) |
|---|---|---|
| *Panel A. Cross-fitted LASSO* | | |
| Age | 0.003 | 0.423 |
| | (0.003) | |
| Interest in history | -0.002* | 0.073 |
| | (0.001) | |
| Self-assessed mathematics ability | -0.002 | 0.199 |
| | (0.001) | |
| Household asset index | 0.002 | 0.339 |
| | (0.002) | |
| *Panel B. Generalized random forest* | | |
| Classroom female share | -0.032*** | $<$ 0.001 |
| | (0.003) | |
| Interest in mathematics | 0.017*** | $<$ 0.001 |
| | (0.002) | |
| Female student | -0.017*** | $<$ 0.001 |
| | (0.003) | |
| Interest in economics | 0.016*** | $<$ 0.001 |
| | (0.002) | |

*Notes:* The LASSO uses school-level folds, while the native random-forest ranking uses observation-level folds. For the gender-stereotypes outcome family, entries describe associations between predetermined characteristics and out-of-fold, model-implied conditional average treatment-effect (CATE) scores. Each row regresses the predicted score on one standardized characteristic, with school-clustered standard errors. A positive gradient means that a one-standard-deviation increase in the characteristic is associated with a larger predicted effect. Within each learner, the four largest absolute gradients among substantively interpretable features are shown; block indicators and other gradients remain in the machine-readable output. The $p$-values test associations with predicted scores, not randomized treatment-by-characteristic interactions. The profiles describe how the fitted learners sort students and should be read alongside the calibration and group-effect tests in Table S46. Those diagnostics do not consistently validate the predicted ordering, so these associations alone do not establish which students benefit most or identify causal mechanisms. Standard errors 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 S53: Predetermined Characteristics Associated with Model-Implied CATE Scores: Economics Applications

| Predetermined characteristic | CATE-score gradient (1) | $p$-value (2) |
|---|---|---|
| *Panel A. Cross-fitted LASSO* | | |
| Age | 0.046* | 0.064 |
| | (0.024) | |
| Classroom female share | 0.044 | 0.143 |
| | (0.029) | |
| Classroom roster size | -0.013 | 0.718 |
| | (0.035) | |
| Interest in mathematics | -0.013* | 0.079 |
| | (0.007) | |
| *Panel B. Generalized random forest* | | |
| Baseline economics application probability | 0.584*** | < 0.001 |
| | (0.051) | |
| Baseline intentions index | 0.548*** | < 0.001 |
| | (0.059) | |
| Baseline identity index | 0.520*** | < 0.001 |
| | (0.051) | |
| Interest in economics | 0.436*** | < 0.001 |
| | (0.040) | |

*Notes:* The LASSO uses school-level folds, while the native random-forest ranking uses observation-level folds. For the economics-application outcome family, entries describe associations between predetermined characteristics and out-of-fold, model-implied conditional average treatment-effect (CATE) scores. Each row regresses the predicted score on one standardized characteristic, with school-clustered standard errors. A positive gradient means that a one-standard-deviation increase in the characteristic is associated with a larger predicted effect. Within each learner, the four largest absolute gradients among substantively interpretable features are shown; block indicators and other gradients remain in the machine-readable output. The $p$-values test associations with predicted scores, not randomized treatment-by-characteristic interactions. The profiles describe how the fitted learners sort students and should be read alongside the calibration and group-effect tests in Table S47. Those diagnostics do not consistently validate the predicted ordering, so these associations alone do not establish which students benefit most or identify causal mechanisms. Standard errors 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 S54: Survey Response by Gender

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Online-endline response* | | | | |
| Spillover vs. control | 1.938 | 8.527** | 6.589* | 0.075 [0.128] |
| | (3.468) | (3.242) | (3.563) | |
| Treated-school assignment | 0.841 | 5.338* | 4.496 | 0.167 [0.211] |
| | (3.128) | (2.823) | (3.164) | |
| Direct treatment vs. control | -0.292 | 1.688 | 1.981 | 0.588 [0.602] |
| | (3.418) | (3.082) | (3.610) | |
| Direct treatment vs. spillover | -2.354 | -6.748* | -4.394 | 0.256 [0.605] |
| | (3.225) | (3.598) | (3.843) | |
| *Panel B. Follow-up-survey response* | | | | |
| Spillover vs. control | 4.081 | 19.274*** | 15.193*** | 0.004 [< 0.001] |
| | (7.370) | (6.830) | (4.814) | |
| Treated-school assignment | 7.084 | 20.304*** | 13.220*** | < 0.001 [< 0.001] |
| | (6.849) | (6.837) | (3.246) | |
| Direct treatment vs. control | 9.992 | 21.544*** | 11.553** | 0.030 [0.095] |
| | (8.763) | (7.295) | (5.032) | |
| Direct treatment vs. spillover | 4.877 | 0.210 | -4.667 | 0.402 [0.605] |
| | (6.224) | (4.627) | (5.540) | |

*Continued on next page*

Table S54 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [wild $q$-value] (4) |
|---|---|---|---|---|
| *Panel C. Immediate-survey response* | | | | |
| Spillover vs. control | 1.201 | -0.748 | -1.949 | 0.335 [0.347] |
| | (1.851) | (1.981) | (1.987) | |
| Treated-school assignment | 2.037 | 0.126 | -1.911 | 0.191 [0.211] |
| | (1.369) | (1.519) | (1.425) | |
| Direct treatment vs. control | 2.861* | 1.139 | -1.722 | 0.252 [0.415] |
| | (1.468) | (1.683) | (1.472) | |
| Direct treatment vs. spillover | 1.603 | 2.294 | 0.691 | 0.738 [0.749] |
| | (2.045) | (2.117) | (2.057) | |

*Notes:* The dependent variables indicate whether a baseline respondent is observed in the online endline, follow-up survey, or immediate survey. Indicators are multiplied by 100, so effects and standard errors are percentage-point changes in response probabilities, not changes in survey indices. A positive effect indicates a higher probability of response. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. No additional covariates are included. The $p$-value tests the women-minus-men difference; bracketed wild-bootstrap $q$-values use Benjamini–Hochberg adjustment across the three response outcomes separately for each assignment contrast. The denominator is final-year baseline respondents with recorded gender, including those not observed at the subsequent wave. These are diagnostics of realized selection after assignment, not estimates of the effect of attending the intervention. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S55: Gender-Specific Estimates with Inverse-Probability Weighting

| Specification | Men (1) | Women (2) | Women - men (3) | $p$-value [$q$-value] diff. (4) |
|---|---|---|---|---|
| *Panel A. Survey wave: Follow-up; Outcome family: Beliefs about economics* | | | | |
| Stabilized inverse-probability weights | 0.082 | 0.193*** | 0.111 | 0.219 [0.346] |
| | (0.083) | (0.053) | (0.088) | |
| IPW, weights capped at the 99th percentile | 0.101 | 0.188*** | 0.087 | 0.272 [0.362] |
| | (0.075) | (0.053) | (0.078) | |
| *Panel B. Survey wave: Follow-up; Outcome family: Identity and self-concept* | | | | |
| Stabilized inverse-probability weights | -0.022 | 0.095 | 0.117* | 0.076 [0.304] |
| | (0.053) | (0.058) | (0.063) | |
| IPW, weights capped at the 99th percentile | -0.022 | 0.094 | 0.116* | 0.078 [0.311] |
| | (0.053) | (0.058) | (0.063) | |
| *Panel C. Survey wave: Follow-up; Outcome family: Intentions and aspirations* | | | | |
| Stabilized inverse-probability weights | 0.036 | 0.016 | -0.020 | 0.829 [0.829] |
| | (0.071) | (0.050) | (0.091) | |
| IPW, weights capped at the 99th percentile | 0.035 | 0.016 | -0.019 | 0.836 [0.836] |
| | (0.072) | (0.051) | (0.092) | |
| *Panel D. Survey wave: Follow-up; Outcome family: Gender stereotypes* | | | | |
| Stabilized inverse-probability weights | 0.140 | 0.035 | -0.105 | 0.260 [0.346] |
| | (0.094) | (0.062) | (0.091) | |
| IPW, weights capped at the 99th percentile | 0.140 | 0.035 | -0.105 | 0.263 [0.362] |
| | (0.094) | (0.062) | (0.092) | |

*Continued on next page*

Table S55 continued

| Specification | Men (1) | Women (2) | Women - men (3) | $p$-value [$q$-value] diff. (4) |
|---|---|---|---|---|
| *Panel E. Survey wave: Immediate; Outcome family: Beliefs about economics* | | | | |
| Stabilized inverse-probability weights | 0.250*** | 0.341*** | 0.091 | 0.176 [0.353] |
| | (0.067) | (0.066) | (0.066) | |
| IPW, weights capped at the 99th percentile | 0.251*** | 0.342*** | 0.091 | 0.174 [0.347] |
| | (0.067) | (0.066) | (0.065) | |
| *Panel F. Survey wave: Immediate; Outcome family: Identity and self-concept* | | | | |
| Stabilized inverse-probability weights | 0.020 | 0.149*** | 0.128** | 0.021 [0.083] |
| | (0.043) | (0.029) | (0.052) | |
| IPW, weights capped at the 99th percentile | 0.016 | 0.154*** | 0.138** | 0.013 [0.051] |
| | (0.043) | (0.030) | (0.052) | |
| *Panel G. Survey wave: Immediate; Outcome family: Intentions and aspirations* | | | | |
| Stabilized inverse-probability weights | -0.027 | -0.037 | -0.010 | 0.886 [0.886] |
| | (0.087) | (0.060) | (0.066) | |
| IPW, weights capped at the 99th percentile | -0.028 | -0.036 | -0.008 | 0.902 [0.902] |
| | (0.087) | (0.060) | (0.066) | |
| *Panel H. Survey wave: Immediate; Outcome family: Gender stereotypes* | | | | |
| Stabilized inverse-probability weights | 0.066 | -0.014 | -0.080 | 0.292 [0.390] |
| | (0.063) | (0.040) | (0.074) | |
| IPW, weights capped at the 99th percentile | 0.062 | -0.013 | -0.076 | 0.307 [0.409] |
| | (0.062) | (0.040) | (0.072) | |

*Notes:* The outcomes are the four follow-up and four immediate survey indices, reported in baseline control-group standard deviations. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Men, Women, and Women - men come from one weighted assignment-by-gender interaction regression with randomization-block fixed effects, the corresponding baseline index, and its missing-value indicator. Standard errors are clustered by school. Response is defined separately for each index and wave as an observed value of that index. Response probabilities are estimated separately by gender using logistic models. The stabilized weight divides the probability conditional on assignment and blocks by the probability additionally conditional on the baseline index, age, GPA, employment, parental university education, interest in economics, and mathematics self-assessment. Missing baseline covariates are mean-imputed and accompanied by missing-value indicators. The capped specification limits these weights to the gender-specific 99th percentile; it does not remove observations above that cutoff. The $p$-value tests the gender difference; bracketed $q$-values apply Benjamini–Hochberg adjustment across four indices within each wave and weighting specification. Response models use final-year baseline respondents; outcome models use observed index responses with valid weights and recorded gender. These estimates assess sensitivity to selection captured by the response models and their baseline covariates. Weighting does not address selection on unobserved outcome determinants. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S56: Placebo Outcomes and Subject-Specific Benchmarks by Gender

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [$q$-value] (4) |
|---|---|---|---|---|
| *Panel A. Interest in English (unrelated-subject placebo)* | | | | |
| Spillover vs. control | -0.015 | 0.036 | 0.051 | 0.451 [0.451] |
| | (0.053) | (0.061) | (0.066) | |
| Treated-school assignment | 0.000 | -0.018 | -0.018 | 0.738 [0.833] |
| | (0.043) | (0.049) | (0.054) | |
| Direct treatment vs. control | 0.009 | -0.082 | -0.090 | 0.133 [0.199] |
| | (0.058) | (0.055) | (0.058) | |
| Direct treatment vs. spillover | 0.075 | -0.104 | -0.179* | 0.084 [0.252] |
| | (0.068) | (0.064) | (0.102) | |
| *Panel B. Interest in history (unrelated-subject placebo)* | | | | |
| Spillover vs. control | 0.052 | -0.065 | -0.116 | 0.134 [0.403] |
| | (0.062) | (0.050) | (0.075) | |
| Treated-school assignment | 0.044 | -0.091** | -0.135** | 0.023 [0.070] |
| | (0.049) | (0.042) | (0.056) | |
| Direct treatment vs. control | 0.035 | -0.124** | -0.159** | 0.017 [0.051] |
| | (0.056) | (0.056) | (0.062) | |
| Direct treatment vs. spillover | 0.007 | -0.038 | -0.045 | 0.654 [0.654] |
| | (0.074) | (0.059) | (0.100) | |
| *Panel C. Interest in Spanish (unrelated-subject placebo)* | | | | |
| Spillover vs. control | -0.089 | -0.017 | 0.072 | 0.421 [0.451] |
| | (0.056) | (0.058) | (0.088) | |
| Treated-school assignment | -0.036 | -0.049 | -0.013 | 0.833 [0.833] |
| | (0.040) | (0.045) | (0.061) | |
| Direct treatment vs. control | 0.010 | -0.082 | -0.092 | 0.286 [0.286] |
| | (0.055) | (0.060) | (0.085) | |
| Direct treatment vs. spillover | 0.096 | -0.048 | -0.144 | 0.220 [0.331] |
| | (0.086) | (0.067) | (0.117) | |
| *Panel D. Self-assessed mathematics ability (related outcome)* | | | | |
| Spillover vs. control | 0.026 | 0.031 | 0.005 | 0.939 [0.939] |
| | (0.059) | (0.073) | (0.069) | |
| Treated-school assignment | -0.003 | 0.047 | 0.050 | 0.454 [0.454] |
| | (0.065) | (0.068) | (0.066) | |
| Direct treatment vs. control | -0.028 | 0.063 | 0.091 | 0.234 [0.467] |
| | (0.084) | (0.077) | (0.075) | |
| Direct treatment vs. spillover | 0.005 | 0.065 | 0.060 | 0.430 [0.430] |
| | (0.065) | (0.054) | (0.075) | |

*Continued on next page*

Table S56 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [$q$-value] (4) |
|---|---|---|---|---|
| *Panel E. Interest in mathematics (related outcome)* | | | | |
| Spillover vs. control | -0.055 | 0.022 | 0.077 | 0.220 [0.440] |
| | (0.049) | (0.045) | (0.062) | |
| Treated-school assignment | -0.042 | 0.012 | 0.055 | 0.337 [0.454] |
| | (0.044) | (0.041) | (0.056) | |
| Direct treatment vs. control | -0.032 | 0.001 | 0.033 | 0.602 [0.602] |
| | (0.054) | (0.045) | (0.063) | |
| Direct treatment vs. spillover | 0.074 | 0.010 | -0.064 | 0.409 [0.430] |
| | (0.058) | (0.044) | (0.077) | |
| *Panel F. Self-assessed economics ability (targeted outcome)* | | | | |
| Spillover vs. control | 0.076 | -0.037 | -0.114 | 0.243 [0.432] |
| | (0.068) | (0.080) | (0.095) | |
| Treated-school assignment | 0.046 | -0.023 | -0.068 | 0.406 [0.478] |
| | (0.067) | (0.070) | (0.081) | |
| Direct treatment vs. control | 0.019 | -0.008 | -0.027 | 0.799 [0.799] |
| | (0.086) | (0.086) | (0.104) | |
| Direct treatment vs. spillover | 0.006 | 0.043 | 0.038 | 0.685 [0.968] |
| | (0.078) | (0.073) | (0.092) | |

*Continued on next page*

Table S56 continued

| Assignment contrast | Men (1) | Women (2) | Women - men (3) | $p$-value [$q$-value] (4) |
|---|---|---|---|---|
| *Panel G. Interest in economics (targeted outcome)* | | | | |
| Spillover vs. control | 0.013 | -0.043 | -0.055 | 0.432 [0.432] |
| | (0.055) | (0.068) | (0.069) | |
| Treated-school assignment | 0.061 | 0.015 | -0.047 | 0.478 [0.478] |
| | (0.041) | (0.056) | (0.065) | |
| Direct treatment vs. control | 0.110** | 0.086 | -0.024 | 0.796 [0.799] |
| | (0.053) | (0.077) | (0.092) | |
| Direct treatment vs. spillover | 0.139* | 0.143** | 0.004 | 0.968 [0.968] |
| | (0.083) | (0.069) | (0.110) | |

*Notes:* The dependent variables are measured at follow-up. Panels A–C report interest in English, history, and Spanish, the unrelated-subject placebo outcomes. Panels D–E report self-assessed ability and interest in mathematics, a related subject; Panels F–G report the corresponding economics outcomes targeted by the intervention. Self-assessed ability records perceived competence, not performance on an objective skills test. Mathematics and economics outcomes may be affected by the intervention and are not placebo tests. Each response is standardized using its baseline control-group mean and standard deviation, with higher values denoting greater interest or self-assessed ability. Men and Women report treatment effects obtained as linear combinations from one pooled assignment-by-gender interaction model. Women - men is their formal difference, not a difference between significance levels. Treated-school assignment compares all classrooms in treated schools, including directly treated and spillover classrooms, with control schools. Direct treatment versus control and spillover versus control compare direct-treatment and spillover classrooms in treated schools, respectively, with control schools. These cross-school models include randomization-block fixed effects and school-clustered standard errors. Direct treatment versus spillover uses only treated schools, school fixed effects, and classroom-clustered standard errors. Regressions also control for the same baseline response. The $p$-value tests the women-minus-men difference. Bracketed $q$-values apply the Benjamini–Hochberg adjustment to conventional clustered $p$-values separately within each assignment contrast and outcome family: the three unrelated-subject placebos, the two mathematics outcomes, and the two economics outcomes. These are not wild-bootstrap values or a joint adjustment across all 28 comparisons. The sample comprises final-year students in the 2026 baseline sample with recorded gender and the variables required for the corresponding regression. The estimation sample can therefore differ across outcomes. These gender-specific analyses are exploratory extensions of the study. The placebo outcomes assess generalized changes in reported subject interest. Failure to reject their null hypotheses does not establish zero effects or rule out selective reporting at either survey horizon. The data come from the study's baseline and subsequent student surveys. Standard errors 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 S57: School-ITT Inference under Alternative Clustering

| Sample | Outcome | Spec. | Estimate | SE school | $p$-value school | SE classroom | $p$-value classroom | Sample size | Schools | Classrooms |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| *Panel A. Sample: Women in the final semester; Outcome: Combined application measure* | | | | | | | | | | |
| Women in the final semester | Combined application measure | PAP specification | 0.333 | 0.318 | 0.304 | 0.353 | 0.347 | 2,009 | 28 | 158 |
| Women in the final semester | Combined application measure | Covariate-adjusted | 0.340 | 0.329 | 0.310 | 0.360 | 0.346 | 2,009 | 28 | 158 |
| *Panel B. Sample: Women in the final semester; Outcome: Administrative applications* | | | | | | | | | | |
| Women in the final semester | Administrative applications | PAP specification | 0.404 | 0.271 | 0.148 | 0.306 | 0.188 | 2,009 | 28 | 158 |
| Women in the final semester | Administrative applications | Covariate-adjusted | 0.432 | 0.275 | 0.128 | 0.303 | 0.156 | 2,009 | 28 | 158 |
| *Panel C. Sample: Women in the final semester; Outcome: UANL economics application* | | | | | | | | | | |
| Women in the final semester | UANL economics application | PAP specification | 0.404 | 0.271 | 0.148 | 0.306 | 0.188 | 2,009 | 28 | 158 |
| Women in the final semester | UANL economics application | Covariate-adjusted | 0.432 | 0.275 | 0.128 | 0.303 | 0.156 | 2,009 | 28 | 158 |
| *Panel D. Sample: Women in the final semester; Outcome: Tec economics (LEC) application* | | | | | | | | | | |
| Women in the final semester | Tec economics (LEC) application | PAP specification | 0.031 | 0.036 | 0.401 | 0.035 | 0.372 | 2,009 | 28 | 158 |
| Women in the final semester | Tec economics (LEC) application | Covariate-adjusted | 0.024 | 0.033 | 0.475 | 0.029 | 0.409 | 2,009 | 28 | 158 |
| *Panel E. Sample: Women in the final semester; Outcome: Economics outside both registries* | | | | | | | | | | |
| Women in the final semester | Economics outside both registries | PAP specification | 0.031 | 0.036 | 0.401 | 0.035 | 0.372 | 2,009 | 28 | 158 |
| Women in the final semester | Economics outside both registries | Covariate-adjusted | 0.024 | 0.033 | 0.475 | 0.029 | 0.409 | 2,009 | 28 | 158 |
| *Panel F. Sample: Women in the final semester; Outcome: Economics application: broad gateway* | | | | | | | | | | |
| Women in the final semester | Economics application: broad gateway | PAP specification | 0.542 | 0.367 | 0.151 | 0.418 | 0.197 | 2,009 | 28 | 158 |
| Women in the final semester | Economics application: broad gateway | Covariate-adjusted | 0.542 | 0.362 | 0.145 | 0.416 | 0.194 | 2,009 | 28 | 158 |
| *Panel G. Sample: Women in the final semester; Outcome: Tec ECSG gateway application* | | | | | | | | | | |
| Women in the final semester | Tec ECSG gateway application | PAP specification | 0.169 | 0.206 | 0.420 | 0.296 | 0.569 | 2,009 | 28 | 158 |
| Women in the final semester | Tec ECSG gateway application | Covariate-adjusted | 0.134 | 0.186 | 0.475 | 0.287 | 0.640 | 2,009 | 28 | 158 |

*Continued on next page*

Table S57 continued

| Sample | Outcome | Spec. | Estimate | SE school | $p$-value school | SE classroom | $p$-value classroom | Sample size | Schools | Classrooms |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| *Panel H. Sample: Women in the final semester; Outcome: Tec application: any program* | | | | | | | | | | |
| Women in the final semester | Tec application: any program | PAP specification | 0.576 | 0.982 | 0.563 | 0.994 | 0.563 | 2,009 | 28 | 158 |
| Women in the final semester | Tec application: any program | Covariate-adjusted | 0.390 | 0.856 | 0.653 | 0.931 | 0.676 | 2,009 | 28 | 158 |
| *Panel I. Sample: Women in the final semester; Outcome: Tec admission: any program* | | | | | | | | | | |
| Women in the final semester | Tec admission: any program | PAP specification | 1.028 | 0.983 | 0.305 | 0.975 | 0.293 | 2,009 | 28 | 158 |
| Women in the final semester | Tec admission: any program | Covariate-adjusted | 0.829 | 0.878 | 0.353 | 0.926 | 0.372 | 2,009 | 28 | 158 |
| *Panel J. Sample: Women in the final semester; Outcome: Tec enrollment: any program* | | | | | | | | | | |
| Women in the final semester | Tec enrollment: any program | PAP specification | 1.008 | 0.874 | 0.259 | 0.907 | 0.268 | 2,009 | 28 | 158 |
| Women in the final semester | Tec enrollment: any program | Covariate-adjusted | 0.828 | 0.799 | 0.310 | 0.875 | 0.345 | 2,009 | 28 | 158 |
| *Panel K. Sample: Women in the final semester; Outcome: Enrollment in Tec ECSG gateway* | | | | | | | | | | |
| Women in the final semester | Enrollment in Tec ECSG gateway | PAP specification | 0.107 | 0.185 | 0.567 | 0.282 | 0.705 | 2,009 | 28 | 158 |
| Women in the final semester | Enrollment in Tec ECSG gateway | Covariate-adjusted | 0.091 | 0.170 | 0.597 | 0.277 | 0.743 | 2,009 | 28 | 158 |
| *Panel L. Sample: Women in the final semester; Outcome: UANL economics exam presented* | | | | | | | | | | |
| Women in the final semester | UANL economics exam presented | PAP specification | 0.404 | 0.271 | 0.148 | 0.306 | 0.188 | 2,009 | 28 | 158 |
| Women in the final semester | UANL economics exam presented | Covariate-adjusted | 0.432 | 0.275 | 0.128 | 0.303 | 0.156 | 2,009 | 28 | 158 |
| *Panel M. Sample: Women in the final semester; Outcome: UANL economics acceptance* | | | | | | | | | | |
| Women in the final semester | UANL economics acceptance | PAP specification | 0.348 | 0.230 | 0.141 | 0.271 | 0.201 | 2,009 | 28 | 158 |
| Women in the final semester | UANL economics acceptance | Covariate-adjusted | 0.360 | 0.236 | 0.139 | 0.274 | 0.191 | 2,009 | 28 | 158 |
| *Panel N. Sample: Women in the final semester; Outcome: Enrollment in an economics degree* | | | | | | | | | | |
| Women in the final semester | Enrollment in an economics degree | PAP specification | -0.014 | 0.169 | 0.934 | 0.199 | 0.944 | 2,009 | 28 | 158 |
| Women in the final semester | Enrollment in an economics degree | Covariate-adjusted | -0.014 | 0.171 | 0.935 | 0.206 | 0.946 | 2,009 | 28 | 158 |
| *Panel O. Sample: All final-semester students; Outcome: Combined application measure* | | | | | | | | | | |
| All final-semester students | Combined application measure | PAP specification | 0.490 | 0.387 | 0.216 | 0.378 | 0.196 | 3,763 | 28 | 159 |

*Continued on next page*

Table S57 continued

| Sample | Outcome | Spec. | Estimate | SE school | $p$-value school | SE classroom | $p$-value classroom | Sample size | Schools | Classrooms |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| All final-semester students | Combined application measure | Covariate-adjusted | 0.614 | 0.371 | 0.109 | 0.397 | 0.124 | 3,763 | 28 | 159 |
| *Panel P. Sample: All final-semester students; Outcome: Administrative applications* | | | | | | | | | | |
| All final-semester students | Administrative applications | PAP specification | 0.412 | 0.325 | 0.215 | 0.317 | 0.196 | 3,763 | 28 | 159 |
| All final-semester students | Administrative applications | Covariate-adjusted | 0.516 | 0.309 | 0.107 | 0.331 | 0.122 | 3,763 | 28 | 159 |
| *Panel Q. Sample: All final-semester students; Outcome: UANL economics application* | | | | | | | | | | |
| All final-semester students | UANL economics application | PAP specification | 0.659 | 0.254 | 0.015 | 0.284 | 0.021 | 3,763 | 28 | 159 |
| All final-semester students | UANL economics application | Covariate-adjusted | 0.741 | 0.251 | 0.006 | 0.303 | 0.015 | 3,763 | 28 | 159 |
| *Panel R. Sample: All final-semester students; Outcome: Tec economics (LEC) application* | | | | | | | | | | |
| All final-semester students | Tec economics (LEC) application | PAP specification | -0.193 | 0.164 | 0.250 | 0.141 | 0.173 | 3,763 | 28 | 159 |
| All final-semester students | Tec economics (LEC) application | Covariate-adjusted | -0.166 | 0.150 | 0.280 | 0.135 | 0.221 | 3,763 | 28 | 159 |
| *Panel S. Sample: All final-semester students; Outcome: Economics outside both registries* | | | | | | | | | | |
| All final-semester students | Economics outside both registries | PAP specification | 0.096 | 0.058 | 0.110 | 0.067 | 0.150 | 3,763 | 28 | 159 |
| All final-semester students | Economics outside both registries | Covariate-adjusted | 0.112 | 0.066 | 0.099 | 0.076 | 0.140 | 3,763 | 28 | 159 |
| *Panel T. Sample: All final-semester students; Outcome: Economics application: broad gateway* | | | | | | | | | | |
| All final-semester students | Economics application: broad gateway | PAP specification | 0.321 | 0.398 | 0.427 | 0.336 | 0.340 | 3,763 | 28 | 159 |
| All final-semester students | Economics application: broad gateway | Covariate-adjusted | 0.432 | 0.385 | 0.272 | 0.354 | 0.223 | 3,763 | 28 | 159 |
| *Panel U. Sample: All final-semester students; Outcome: Tec ECSG gateway application* | | | | | | | | | | |
| All final-semester students | Tec ECSG gateway application | PAP specification | -0.284 | 0.280 | 0.319 | 0.184 | 0.123 | 3,763 | 28 | 159 |
| All final-semester students | Tec ECSG gateway application | Covariate-adjusted | -0.249 | 0.268 | 0.362 | 0.182 | 0.173 | 3,763 | 28 | 159 |
| *Panel V. Sample: All final-semester students; Outcome: Tec application: any program* | | | | | | | | | | |
| All final-semester students | Tec application: any program | PAP specification | 0.779 | 0.980 | 0.434 | 0.681 | 0.254 | 3,763 | 28 | 159 |

*Continued on next page*

Table S57 continued

| Sample | Outcome | Spec. | Estimate | SE school | $p$-value school | SE classroom | $p$-value classroom | Sample size | Schools | Classrooms |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| All final-semester students | Tec application: any program | Covariate-adjusted | 0.670 | 0.819 | 0.420 | 0.640 | 0.297 | 3,763 | 28 | 159 |
| *Panel W. Sample: All final-semester students; Outcome: Tec admission: any program* | | | | | | | | | | |
| All final-semester students | Tec admission: any program | PAP specification | 0.924 | 0.758 | 0.234 | 0.667 | 0.168 | 3,763 | 28 | 159 |
| All final-semester students | Tec admission: any program | Covariate-adjusted | 0.804 | 0.622 | 0.207 | 0.638 | 0.209 | 3,763 | 28 | 159 |
| *Panel X. Sample: All final-semester students; Outcome: Tec enrollment: any program* | | | | | | | | | | |
| All final-semester students | Tec enrollment: any program | PAP specification | 0.887 | 0.670 | 0.197 | 0.693 | 0.203 | 3,763 | 28 | 159 |
| All final-semester students | Tec enrollment: any program | Covariate-adjusted | 0.800 | 0.624 | 0.211 | 0.690 | 0.248 | 3,763 | 28 | 159 |
| *Panel Y. Sample: All final-semester students; Outcome: Enrollment in Tec ECSG gateway* | | | | | | | | | | |
| All final-semester students | Enrollment in Tec ECSG gateway | PAP specification | -0.107 | 0.150 | 0.482 | 0.149 | 0.473 | 3,763 | 28 | 159 |
| All final-semester students | Enrollment in Tec ECSG gateway | Covariate-adjusted | -0.083 | 0.140 | 0.557 | 0.145 | 0.567 | 3,763 | 28 | 159 |
| *Panel Z. Sample: All final-semester students; Outcome: UANL economics exam presented* | | | | | | | | | | |
| All final-semester students | UANL economics exam presented | PAP specification | 0.595 | 0.258 | 0.029 | 0.280 | 0.035 | 3,763 | 28 | 159 |
| All final-semester students | UANL economics exam presented | Covariate-adjusted | 0.669 | 0.254 | 0.014 | 0.297 | 0.026 | 3,763 | 28 | 159 |
| *Panel AA. Sample: All final-semester students; Outcome: UANL economics acceptance* | | | | | | | | | | |
| All final-semester students | UANL economics acceptance | PAP specification | 0.346 | 0.183 | 0.070 | 0.218 | 0.114 | 3,763 | 28 | 159 |
| All final-semester students | UANL economics acceptance | Covariate-adjusted | 0.379 | 0.173 | 0.037 | 0.227 | 0.097 | 3,763 | 28 | 159 |
| *Panel AB. Sample: All final-semester students; Outcome: Enrollment in an economics degree* | | | | | | | | | | |
| All final-semester students | Enrollment in an economics degree | PAP specification | -0.124 | 0.216 | 0.570 | 0.230 | 0.590 | 3,763 | 28 | 159 |
| All final-semester students | Enrollment in an economics degree | Covariate-adjusted | -0.056 | 0.212 | 0.793 | 0.234 | 0.810 | 3,763 | 28 | 159 |

*Notes:* Panels distinguish the women-only and all-student final-semester samples and the displayed application definitions. Estimate is the same treated-school assignment coefficient under both inference columns. The School columns cluster standard errors at the first-stage assignment unit and provide the design-aligned primary inference. The Classroom columns hold the coefficient and specification fixed but cluster at the finer classroom level. Every model includes the original randomization-block fixed effects and the adjustment set named in Spec. Application outcomes are binary indicators multiplied by 100; effects, standard errors, and control means are therefore reported in percentage points. Classroom clustering is a sensitivity comparison; it does not replace school clustering when assignment occurs at the school unit. Significance is assessed using two-sided, unrounded $p$-values below 0.10.