EdWorkingPaper No. 26-1498

First Impressions Matter: Instructor Gender and Women's Persistence in Economics

Marissa Eckrote-Nordland
University of Wisconsin–La Crosse

Lisa Giddings
University of Wisconsin–La Crosse

Laurie Miller
University of Wisconsin–La Crosse

John M. Nunley
University of Wisconsin–La Crosse

Garrett Soper
University of Wisconsin–La Crosse

Using near-random assignment of students to instructors in introductory economics at a broad-access public university, we study how instructor gender affects women's persistence in economics. Female instructors close roughly 40 percent of the gender gap in advanced economics course-taking, with a similar but less precisely estimated improvement in major completion. The effect is concentrated in the first introductory course: exposure to a female instructor in the second course adds no detectable persistence benefit once the first instructor's gender is accounted for. Female instructors also improve women's relative performance in introductory economics, but grades explain no more than about one-third of the gender-match effect. Class fixed effects and the absence of a detectable male-student response make it difficult to attribute the effect to general instructor quality. The remaining effect is consistent with mechanisms related to belonging, representation, identity, role modeling, or perceived fit. The findings point to introductory-course staffing as a practical tool for shaping women's persistence in economics.

VERSION: June 2026

Suggested citation: Eckrote-Nordland, Marissa, Lisa Giddings, Laurie Miller, John M. Nunley, and Garrett Soper. (2026). First Impressions Matter: Instructor Gender and Women's Persistence in Economics. (EdWorkingPaper: 26-1498). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/5bbj-0v49

First Impressions Matter: Instructor Gender and Women's Persistence in Economics *

Marissa Eckrote-Nordland,† Lisa Giddings,‡ Laurie Miller,§

John M. Nunley,¶ and Garrett Soper ||

June 8, 2026

Abstract

Using near-random assignment of students to instructors in introductory economics at a broad-access public university, we study how instructor gender affects women's persistence in economics. Female instructors close roughly 40 percent of the gender gap in advanced economics course-taking, with a similar but less precisely estimated improvement in major completion. The effect is concentrated in the first introductory course: exposure to a female instructor in the second course adds no detectable persistence benefit once the first instructor's gender is accounted for. Female instructors also improve women's relative performance in introductory economics, but grades explain no more than about one-third of the gender-match effect. Class fixed effects and the absence of a detectable male-student response make it difficult to attribute the effect to general instructor quality. The remaining effect is consistent with mechanisms related to belonging, representation, identity, role modeling, or perceived fit. The findings point to introductory-course staffing as a practical tool for shaping women's persistence in economics.

JEL: I23, I26, J16, J24, A22

Keywords: gender, economics education, major choice, role models, human capital

*We thank Riley Acton, Duha Altindag, Sandra Black, Taggert Brooks, David Card, Todd Elder, Graciela Engen, James Murray, R. Alan Seals, Amir Tayebi, Nicole Vidden, the Institutional Review Board and Office of Institutional Research, Assessment, and Planning at the University of Wisconsin–La Crosse, and participants at the 2026 Society of Labor Economists, Eastern Economic Association, and CTREE annual meetings. Errors are our own.

†Eckrote-Nordland, Assistant Professor, Department of Economics, University of Wisconsin–La Crosse, meckrote@uwlax.edu.

‡Giddings, Professor, Department of Economics, University of Wisconsin–La Crosse, lgiddings@uwlax.edu.

§Miller, Teaching Professor, Department of Economics, University of Wisconsin–La Crosse, lmiller2@uwlax.edu.

¶Nunley, Professor, Department of Economics, University of Wisconsin–La Crosse, jnunley@uwlax.edu.

||Soper, Institute Policy Analyst I, Institutional Research, Assessment, and Planning, University of Wisconsin–La Crosse, vsoper@uwlax.edu.

Data Availability Statement: Our analysis relies on confidential student-record data governed by the Family Educational Rights and Privacy Act (FERPA). Because the data contain identifiable educational records, they cannot be publicly released. However, independent researchers may obtain access by entering into a data-use agreement with our university. We would be pleased to assist researchers in facilitating such access, and we will provide full replication code and documentation.

Disclosure Statement: We declare that we have no relevant or material financial interests that relate to the research described in this paper. We received no funding to support this research.

Institutional Review Board (IRB): This research was reviewed and approved by the Institutional Review Board (IRB) of the University of Wisconsin–La Crosse under expedited review procedures (Protocol ID: IRB_25-MEN-179). The study uses deidentified administrative data provided by the university, and no direct interaction with human subjects occurred. All procedures complied with applicable ethical standards for research involving human subjects.

1 Introduction

Women have made considerable gains across much of science, though they remain underrepresented in fields such as engineering and computer science (National Center for Science and Engineering Statistics, 2021). Economics is another quantitatively intensive field where progress has stalled. Women earn the majority of U.S. bachelor's degrees but only 30 to 35 percent of economics degrees, with comparable shares earning economics PhDs and holding economics faculty positions (Chevalier, 2021; Bayer and Rouse, 2016; Board of Governors of the Federal Reserve System, 2021; Lundberg and Stearns, 2019). The stakes extend beyond the discipline itself. Economics graduates are 2.8 times as likely as other bachelor's holders to work in finance and insurance and 2.3 times as likely to work in occupations that shape regulation, public administration, and economic analysis.[1] Because male and female economists hold different views on policy questions (May et al., 2021), the gender composition of the profession may influence the views that inform those decisions.

Introductory courses are a natural place to look for the roots of these gaps. Many women who earn STEM bachelor's degrees did not initially plan to major in a STEM field, suggesting that early college coursework can shape entry into quantitatively intensive fields (Xie and Shauman, 2004). A growing literature documents that instructor gender, role models, and representation can affect women's persistence in fields where women remain underrepresented. What remains less clear is whether these effects emerge at a student's first encounter or accumulate through repeated exposure, whether they operate mainly through grades or through non-grade channels such as belonging, identity, representation, and perceived fit, and whether the same patterns appear outside the selective, well-resourced institutions that dominate much of the existing causal evidence. This paper asks whether early exposure to female instructors can change who continues in economics, and when and why that exposure matters.

We test whether instructor-gender effects on women's persistence come from a student's first encounter or build with repeated exposure across the introductory sequence. The two-course structure of introductory economics, taken by roughly 83 percent of our sample, lets us identify both effects. No prior study has been able to make this distinction in any field. This work builds on causal evidence from other higher-education settings documenting that instructor gender can shape women's course performance, persistence, and major choice. At the U.S. Air Force Academy, random assignment to female instructors closes the gender gap in STEM major choice among high-ability women, with effects that persist after graduation (Carrell et al., 2010; Mansour et al., 2022). Other quasi-random college settings find related instructor-gender effects on grades, dropout rates, and subsequent course-taking, though the effects vary by classroom context and are not always specific to women (Hoffmann and Oreopoulos, 2009; Maurer et al., 2023). In economics, existing evidence comes primarily from interventions that introduce role models or information about the field rather than from exposure to course instructors themselves. Female alumni visits increase women's subsequent economics course-taking and majoring (Porter and Serra, 2020). Same-gender alumni speakers increase economics course-taking for both men and women (Patnaik et al., 2024). Informational

[1] Authors' calculations from the 2009–2019 American Community Survey (IPUMS USA, 2025). The composite includes lawyers, public administrators, management analysts, and economists.

nudges produce more limited or subgroup-specific effects (Chambers et al., 2021; Pugatch and Schroeder, 2021, 2024). Random assignment to female faculty advisors reduces dropout and raises completion among women, but the study tests advisors rather than instructors, and the sample includes only students who have already chosen economics as their major (Canaan and Mouganie, 2021).

Our analysis uses administrative data from the University of Wisconsin–La Crosse (UWL), a mid-sized, predominantly undergraduate public university. The sample consists of 3,018 first-year students placed into introductory microeconomics or macroeconomics between Fall 2009 and Fall 2019, taught by 28 instructors (14 women, 14 men), with the final cohort tracked through Summer 2025. During the study period, first-year students were administratively pre-assigned to introductory microeconomics and macroeconomics sections before registration, while the department chair assigned instructors to introductory sections separately on the basis of teaching preferences and staffing constraints. These administrative features generate near-random variation in exposure to male and female economics instructors in students' first semester. As a regional public university, UWL allows us to extend prior evidence to the broad-access segment of higher education. Regional public universities like these enroll about 70 percent of the seven million undergraduates at U.S. public four-year institutions (American Association of State Colleges and Universities, 2026).

Female introductory instructors reduce the gender gap in advanced economics course-taking by roughly 40 percent. The gap is about 15 percentage points under male instructors and 8 percentage points under female instructors. The effect on economics-major completion is larger in proportional terms (60 to 70 percent) but less precisely estimated, with the gap falling from about four percentage points under male instructors to one and a half percentage points under female instructors. For a typical UWL entering cohort of about 109 women in introductory economics, these estimates imply roughly seven additional women taking an advanced economics course and three additional women completing an economics major when the introductory instructor is female rather than male.

The effect is concentrated in the first introductory course. A female instructor in the second course shows no detectable effect, whether estimated on its own or conditional on the first instructor's gender. The pattern suggests that representation matters at the first encounter rather than through repeated exposure. Female instructors are not simply better teachers across the board, as instructor gender has no detectable effect on male students and specifications that absorb persistent differences across instructors yield the same result. Female instructors give higher grades to all students. The additional boost for women is 0.15 to 0.16 GPA points, but this explains less than one-third of the effect on taking advanced economics courses and majoring in the field. Most of the effect is consistent with channels related to belonging, role modeling, and identity. The results are not driven by a single instructor, a particular subset of the instructor pool, or differences in observable section and instructor characteristics. The estimates also survive alternative inference procedures and sensitivity checks for selection into the second introductory course.

The findings identify introductory-course staffing as a practical tool for shaping the economics pipeline. Because the effect appears in the first introductory course rather than building through later exposure, the initial section assignment is where staffing matters. Unlike mentoring or informational interventions that reach only students who opt in, staffing decisions affect the full entering cohort, including first-generation students and others who might not seek out mentors on their own. The main constraint is the limited supply of female faculty in economics. The estimates therefore speak most directly to departments moving from low female-faculty representation toward greater balance in gateway courses. In departments where women teach only about a third of introductory sections (Lundberg and Stearns, 2019), moving to gender parity in gateway-course staffing would raise women's exposure to a female first instructor by 17 percentage points. The result is an increase in advanced economics course-taking by about 7 percent (about 1.1 additional women per cohort) and economics-major completion by about 17 percent (about 0.4 additional women per cohort).

2 Institutional Setting

Our research setting is the University of Wisconsin–La Crosse (UWL), a mid-sized, predominantly undergraduate public university located in southwestern Wisconsin. Between 2009 and 2019, enrollment rose to close to 10,000 students, a majority of whom were women and the vast majority of whom were white in-state residents. The economics department is housed within the College of Business Administration (CBA), where male faculty outnumber female faculty by about three to two. Students can major in economics through the CBA or the College of Arts, Social Sciences, and Humanities (CASSH), with most majors enrolled through the CBA. All students in the CBA are required to complete both introductory microeconomics and macroeconomics, and most students do so during their first year. Students pursuing economics through CASSH typically take both introductory courses early in their programs as well. The two courses are not sequenced as prerequisites, so a student administratively assigned to microeconomics in the fall typically takes macroeconomics in the spring, and vice versa, with the spring instructor selected through standard self-registration. Roughly 83 percent of students in our analytic sample take both introductory courses. In Section 5.3.1, we use this subsample to ask whether the effect comes from the first female instructor students encounter or builds with later exposure to additional female instructors.

The CBA standardizes content across sections of its gateway courses through common learning outcomes. For each gateway course in the college, including introductory microeconomics and introductory macroeconomics, there is a common set of learning objectives (LOs) and faculty design their courses around these learning outcomes. Additionally, the LOs are assessed through a common exam question embedded in each instructor's exam. Under UWL's textbook rental system, all sections of a given course use a common textbook chosen by the course faculty. Individual instructors retain discretion over homework, quiz, and exam design beyond the embedded common item. The result is that section-level differences in covered material are limited by design, even though pedagogy and assessment style may vary across instructors. As a result, differences in student outcomes across sections cannot be attributed to differences in core content.

During the study period, incoming first-year students enrolled in introductory economics had no opportunity to choose their instructors. The Registrar's Office built schedules administratively and introductory sections were pre-loaded into students' schedules before registration began. Because course sections filled entirely during pre-registration, students rarely changed sections thereafter.[2] Faculty assignments were finalized before pre-registration, ensuring that instructor allocation preceded student placement. The department chair assigns instructors based on teaching preferences, scheduling constraints, and departmental needs, not student composition. These parallel administrative processes yield a near-random assignment of students to instructors, conditional on enrollment timing. Consistent with this process, observable student characteristics are balanced across instructor gender, as confirmed by the balance tests in Section 4.2.

3 Identification Strategy

To estimate the effect of instructor–student gender matching on students' subsequent persistence in economics, we estimate

$$y_{i,c,j,t} = \alpha + \beta D_i + \gamma D_{j(i)} + \delta(D_i \times D_{j(i)}) + X_i'\Theta_1 + X_{c(i)}'\Theta_2 + X_{j(i)}'\Theta_3 + \eta_{s_{\mathcal{P}}(c,t)} + \varepsilon_{i,c,j,t}. \quad (1)$$

The binary variable $y_{i,c,j,t}$ denotes either (i) enrollment in any advanced economics course or (ii) completion of an economics major for student i in class c taught by instructor j in term t. The variables D_i and $D_{j(i)}$ are gender indicators for the student and instructor, respectively, and X_i, $X_{c(i)}$, and $X_{j(i)}$ include student-, class-, and instructor-level covariates listed in Table 1 and defined in Appendix Table A1.[3] We also include controls for the dates that students applied to the university as well as their registration dates, separated into deciles within each term and entered as factor variables in X_i. The term $\varepsilon_{i,c,j,t}$ captures variables that affect $y_{i,c,j,t}$ but are not accounted for among the right-hand-side variables.

The term $\eta_{s_{\mathcal{P}}(c,t)}$ captures fixed effects at different levels of aggregation, where the partition $\mathcal{P}$ determines the scope of within-group comparisons. When $\mathcal{P} = \phi_t$, the fixed effects absorb semester-level shocks (e.g., curricular changes). When $\mathcal{P}$ = *assignment cell*, defined by the interaction of term $\times$ course $\times$ day $\times$ time, comparisons are restricted to sections offered in the same time slot within a term.[4] When $\mathcal{P} = \phi_c$, class (section) fixed effects absorb all instructor and section characteristics, so identification of δ comes solely from within-class differences between male and female students. Because instructor gender is perfectly collinear with the class identifier, the class-fixed-effects specification cannot separately identify the main effect

[2] About 4% of our sample registered for their first-semester introductory economics course after the pre-registration period. In a sensitivity check, we exclude any students whose registration for either introductory economics course occurred in August or September. Results are robust.

[3] Female and male instructors differ along several observable dimensions in our sample. Female instructors are more likely to be foreign-born, less likely to be tenured or on the tenure track, and teach modestly larger sections on average (Appendix Table A2). To ensure that the estimated gender-match coefficient δ does not capture these other instructor attributes rather than instructor gender itself, we include controls for instructor nativity, tenure status, teaching experience, class size, and section gender composition in every specification.

[4] Our sample includes 11 Fall terms, two pre-assigned introductory courses (microeconomics and macroeconomics), three day patterns (Monday/Wednesday/Friday, Tuesday/Thursday, and remote), and five time blocks (morning, mid-day, late afternoon, evening, and remote). Each distinct combination of these categories constitutes a unique assignment cell.

of instructor gender (γ) and instead identifies only the differential response of female versus male students to the same instructor within a section.

We treat the assignment-cell specification as preferred because it aligns directly with the institutional block-assignment process that generates the quasi-random exposure. We additionally estimate specifications with class, instructor, and instructor-by-term fixed effects, which absorb increasingly rich sets of persistent instructor- and section-level determinants of student outcomes (see Section 5.2).

The α, β, γ, δ, and Θ_m are parameters to be estimated. The coefficient β on the female-student indicator D_i is the gender gap in the outcome among students taught by male instructors, and $\beta + \delta$ is the gender gap among students taught by female instructors. The interaction coefficient δ is the difference between these two gaps and, therefore, is the coefficient of interest. Under our identification assumptions, δ can be interpreted as the causal effect of instructor–student gender matching on persistence in economics, particularly with assignment-cell or class fixed effects held constant.

Because students are drawn from common registration pools within assignment cells, we cluster standard errors at that level to allow arbitrary correlation in residuals within shared scheduling environments. We also report standard errors with two-way clustering by assignment cell and instructor to account for serial correlation in instructor shocks (e.g., see Cameron et al., 2011; Cameron and Miller, 2015). Results are nearly identical. Because the instructor cluster dimension is small (roughly 28 unique instructors), the asymptotic justification for two-way cluster-robust standard errors is weak. We read the two-way figures as a sensitivity check on the assignment-cell baseline rather than as the primary inference.

4 Data

4.1 Student, Class, and Instructor Characteristics

We draw on 11 years of administrative records covering all introductory economics sections at UWL between 2009 and 2019. The analytic sample is the full set of students administratively placed into introductory microeconomics or macroeconomics over this period.[5] The records include student-level, instructor-level, and course-level fields, and the final cohort (entering Fall 2019) is tracked through Summer 2025, providing six years of follow-up to observe advanced course-taking decisions and graduation outcomes.

Students arrive with varied academic backgrounds (Table 1). Mean ACT composite is 24.4, in line with national averages for four-year public universities. Math placement is distributed across remedial (20%), college algebra (23%), pre-calculus (40%), and calculus (16%). About 29% are first-generation, 22% are Pell-eligible, and 9% identify as students of color. Roughly 88% attended public high schools, and incoming students bring 5.4 college credits on average from Advanced Placement, International Baccalaureate, or dual-enrollment coursework. On average, incoming students enroll in 15 credits during their first semester.

[5] We exclude students who are placed into both courses in their first semester from our analysis. Over the 11-year sample period, around 3% of students were enrolled in both introductory courses simultaneously in their first semester at UWL.

Introductory sections average 33 students, and the average student is taught by an instructor with 13 years of teaching experience. About 54% of students are taught by tenured or tenure-track faculty and 36% by foreign-born instructors. Women constitute 39% of enrolled students, with little variation across terms or instructor gender.

Male and female students enter on broadly similar footing. ACT scores, math placement, and college affiliation differ by less than 0.12 standard deviations or 4 percentage points across the two groups. Women are about 5 percentage points more likely to be first-generation (33% versus 27%) and arrive with roughly two more transfer credits (6.7 versus 4.5), and we include both in the baseline controls. Differences in the share taught by foreign-born or tenured instructors are under 5 percentage points, and average class size is effectively identical.

The gender gap emerges in persistence rather than preparation. About 23% of introductory students take at least one advanced economics course, but the rate is 27% for men and 17% for women. Only about 4% complete an economics major, 4.5% of men versus 2.8% of women. Female instructors taught 50% of students in the analytic sample.

4.2 Covariate Balance

To assess the validity of the research design, we examine whether instructor gender is balanced with respect to student characteristics and enrollment timing. We evaluate covariate balance using standardized differences. Absolute standardized differences below 0.10 are generally considered negligible, while values below 0.25 are commonly treated as adequate balance (Austin, 2009; Rubin, 2001) (Figure 1). Computed within assignment cells, all thirteen covariates fall well within the 0.10 range, with the largest (White versus non-White) equal to 0.05 in absolute value. Student characteristics are therefore well balanced across male and female instructors.

Because section assignment is plausibly quasi-random conditional on enrollment timing and assignment cell, we test whether application and registration dates predict assignment to a female instructor. At the student level, joint Wald tests yield $F = 1.38$ $(p = 0.15)$ for the exclusion of the application and registration date variables. When aggregated to the section level so that each class receives equal weight, the timing variables jointly yield $F = 0.58$ $(p = 0.90)$. Together, these results are consistent with conditional random assignment. Timing variables do not predict instructor gender either at the student level or at the section level, and student characteristics that might otherwise concern us are absorbed by the controls and assignment-cell fixed effects.

We complement the regression-based balance checks with a permutation-based randomization test that reassigns instructor gender within assignment cells while preserving the observed allocation structure. Across the full set of covariates, the results are consistent with conditional random assignment: only one covariate rejects at conventional levels before multiple-testing adjustment, and none do after Westfall-Young correction (see Appendix C).

4.3 External Validity

UWL belongs to a broad cluster of mid-sized, predominantly undergraduate, non-flagship public universities that together educate the majority of U.S. undergraduates who encounter introductory economics. Benchmarking our setting against the U.S. Department of Education's College Scorecard places UWL alongside roughly 100 peer institutions on the dimensions of size, selectivity, affordability, and instructional resources. By contrast, the universities studied in much of the prior gender-match literature (the U.S. Air Force Academy, Southern Methodist University, and the University of Wisconsin–Madison among them) are more selective and substantially better resourced. Our estimates therefore complement rather than substitute for that prior evidence, extending it to the broad-access segment of public higher education where most U.S. undergraduates are first exposed to the field.

Three caveats are worth noting. First, the treatment variation comes from a pool of 14 female and 14 male instructors. Robustness checks (leave-one-out exclusion of each female instructor and partitions of the instructor pool by tenure and teaching experience, Section 5.2) show that the aggregate effect is not driven by any single instructor or subset, but whether UWL's female faculty are stylistically representative of female introductory-economics faculty more broadly cannot be tested directly with these data. Second, UWL's economics department is unusually gender-balanced (14 female, 14 male instructors), so the estimate speaks to policy moves from a low female-faculty baseline toward greater balance, a lever with meaningful room at most institutions even if not at this one. The female faculty who teach gateway courses at a near-parity department like UWL may differ on unobservables from those who would teach gateway courses at departments moving toward balance from a low baseline. This margin is outside what the design can identify. Third, UWL's student body is predominantly White and from the upper Midwest, a demographic context in which perceived similarity between students and U.S.-born female instructors may be unusually high, consistent with the larger effects we observe under U.S.-born instructors. The magnitude may differ in settings with greater demographic heterogeneity. Appendix B reports the peer-institution k-means clustering procedure and the universities clustered with UWL. Table 2 reports the institutional characteristics of UWL (Panel A), the peer cluster (Panel B), and universities featured in the prior gender-match literature (Panel C).

5 Results

5.1 Baseline Estimates

Table 3 presents baseline estimates of gender-match effects on students' continuation in economics. Panel A reports results for the probability of taking at least one advanced economics course. The gender gap is 14.7 to 15.2 percentage points when the instructor is male and 8.2 to 8.4 percentage points when the instructor is female. The 6.2 to 6.8 percentage-point difference implies a gap that is 42 to 45% smaller under female instructors. For a typical entering cohort of about 109 women in introductory economics at UWL, this differential implies roughly 7 additional female advanced-course-takers when the introductory instructor is female rather than male, holding the assignment cell fixed.

Panel B examines the probability of completing an economics major. Female students taught by male instructors are about 3.5 to 4.0 percentage points less likely than men to major in economics, compared with a gap of about 1.3 to 1.4 percentage points among those taught by female instructors. The male-instructor gap is statistically significant at the 1% level. The female-instructor gap is not statistically distinguishable from zero. The interaction implies the gap narrows by 60 to 67% when students are assigned a female instructor. Although less precisely estimated than the advanced-course result, the interaction reaches at least 10% significance under term and assignment-cell fixed effects (columns 1 and 2) and 5% significance under class fixed effects (column 3). The 2.4 percentage-point coefficient on major completion translates to roughly 3 additional female economics majors per cohort within the same comparison.

We do not report the estimated effects of instructor gender on male students in Table 3. Male students' continuation rates are unaffected by whether they were taught by male or female instructors, implying that the benefit of female instructors accrues specifically to female students rather than reflecting a general instructor effect. [6]

In sum, female instructors narrow the gender gap in advanced-course enrollment by nearly half, precisely estimated across specifications. The major-completion estimate implies a proportionally similar reduction but is less precisely estimated under conservative inference. The estimates are stable across all three fixed-effects specifications. The within-class gender-match coefficient (column 3) matches the within-cell coefficient (column 2), so residual multi-cell variation does not drive the result.

5.2 Robustness Checks

We verify the baseline estimate against five sensitivity checks addressing functional form, sample composition, the instructor pool, instructor quality, and cluster-robust inference under a small instructor-cluster dimension.

First, we check whether δ is sensitive to restrictive slope assumptions in the controls. Appendix Tables A5 and A6 add female-student interactions with each of the five instructor- and class-level attributes (foreign-born, tenured, instructor experience, class size, and section share female). The gender-match coefficient is essentially unchanged. Appendix Table A7 adds female-student interactions with each of the eleven baseline student covariates and with the term fixed effect. In these saturated specifications, the lower-order coefficients no longer admit a simple standalone interpretation because the effect of female-student status is allowed to vary flexibly with observed characteristics. The coefficient of interest remains the female-student-by-female-instructor interaction, which continues to capture the differential persistence response of women assigned to female instructors. The female-student-covariate interactions are jointly significant in every specification ($p < 0.001$ for the covariate block under both clustering schemes), indicating that female students respond differently to several baseline covariates. Allowing for this heterogeneity does not attenuate the gender-match estimate. The augmented specification with the full set of female-student-covariate interactions gives a gender-match coefficient of 8.4 percentage points on advanced-course enrollment in the preferred assignment-cell column, compared with 6.3 percentage points in the baseline, an

[6] Appendix Tables A3 and A4 report individual β, γ, and δ estimates from equation 1. The estimate for γ is the predicted outcome difference among male students assigned to female versus male instructors. Class fixed effects fully absorb the instructor-gender indicator (perfect collinearity with class identifiers), so γ cannot be separately identified. The Appendix tables report the column 2 specification.

increase of roughly one-third. The gender-match coefficient for the major outcome rises from 2.4 to 3.9 percentage points across the same comparison, an increase of about 60 percent, and reaches the five-percent significance level under both one- and two-way clustering schemes. We retain the parsimonious baseline as the headline because it is the simpler reduced form, but the comparison suggests that restricting female students to share common slopes with male students attenuates rather than inflates the estimated gender-match effect.

Second, we check that no single female instructor drives the result. Appendix Table A8 reports a leave-one-out analysis that sequentially excludes each of the 14 female instructors and re-estimates the preferred assignment-cell specification. The baseline gender-match coefficient is stable across all 14 jackknife samples, with every jackknife estimate lying within one baseline standard error of the full-sample point estimate for both outcomes.

Third, we check that the result is not driven by a specific subset of the 14 female instructors. Appendix Table A9 reports the gender-match coefficient within partitions of the instructor pool by tenure status and by years of teaching experience at UWL. The coefficient is stable across partitions and the equality test fails to reject ($p = 0.63$ for tenure status, $p = 0.79$ for experience, advanced-course outcome).

Fourth, we check whether the gender-match effect reflects differences in instructor quality or pedagogy between male and female instructors. Prior work suggests that same-gender instructor effects may partly reflect differences in instructional strategies or role-model channels (Andersen and Reimer, 2019; Bettinger and Long, 2005), while broader evidence shows that teacher quality and observable classroom practices can affect student learning (Araujo et al., 2016; Lavy, 2016). Relative to the baseline assignment-cell specification in Table 3, the gender-match estimate remains quantitatively similar in specifications with instructor and instructor-by-term fixed effects (Appendix Tables A3 and A4, columns 2–3). Across these different comparisons, the estimates change little, making it difficult to attribute the results primarily to differences in pedagogy, grading standards, or instructor quality.

Fifth, we address the small effective cluster dimension at the instructor level. With roughly 28 instructor clusters the asymptotic justification for cluster-robust inference is weak (Cameron and Miller, 2015; MacKinnon and Webb, 2017), so we complement the conventional standard errors with a wild cluster bootstrap clustering at the instructor level, imposing the null on the gender-match interaction and using Webb six-point weights with 9,999 replications (Roodman et al., 2019). Appendix Table A10 reports the resulting bootstrap p-values across the three fixed-effects specifications and both outcomes. The advanced-course estimate retains conventional significance under all three specifications, with bootstrap p-values between 0.04 and 0.07. The major-completion estimate is at or just outside the ten-percent level, with bootstrap p-values between 0.07 and 0.10. The bootstrap delivers the same qualitative reading as the conventional standard errors, with the major-completion estimate again less precisely estimated than the advanced-course estimate.

Sixth, we examine whether the gender-match effect varies systematically across student, instructor, and classroom subgroups. Appendix Tables A11 and A12 report interaction estimates by first-generation status, math placement, instructor nativity, classroom gender composition, and course type, under one-way and two-way clustered standard errors respectively.

Appendix Tables A13 and A14 report cross-cutting decompositions of the first-generation contrast for the advanced-course and major-completion outcomes. Point estimates are largely consistent in sign across subgroups. Some cuts show seemingly larger effects, but the subgroup samples are small and the design is underpowered to draw conclusions about heterogeneity.

5.3 Evidence on Mechanisms

5.3.1 First Exposure versus Cumulative Exposure

The timing of the response helps distinguish whether one early exposure is sufficient or whether the effect grows as students encounter additional female instructors. If the effect builds with later exposure, students who have female instructors in both introductory courses should be more likely to continue than students who have a female instructor only in the first course. If the effect operates primarily through first exposure, the second instructor's gender should add little once the first instructor's gender is held constant.

We probe this question on the subsample of students who took both introductory courses (roughly 83% of our analytic sample), adding the gender of the second-semester instructor to the baseline specification. Unlike the first course, students choose their section for the second course during standard registration, so second-instructor gender is not exogenously assigned. A natural concern is that women who responded to a female first instructor could select into female-taught sections the second time around. Table 4 regresses second-instructor gender on the first-instructor gender-match interaction and the full set of controls. Across all three fixed-effects specifications the coefficient is small, negative, and statistically insignificant. This selection channel appears negligible in these data.

Tables 5 and 6 report the augmented gender-match specification on the both-courses subsample, for the advanced-course and economics-major outcomes respectively. Each table pairs a comparison row, re-estimating the baseline gender-match coefficient on this subsample without the second-instructor controls, with an augmented specification that adds the second-semester instructor's gender as a main effect and as an interaction with the female-student indicator. The pattern is the same in every column. The first-instructor gender-match coefficient differs from the comparison-row estimate by less than two-tenths of a percentage point in every fixed-effects specification. The second-instructor main effect and its interaction with female student are individually small and statistically insignificant, and the joint Wald test that the entire second-instructor block equals zero fails to reject in every column under both clustering schemes, with p-values ranging from 0.70 to 0.98. The class-fixed-effects specification (column 3) absorbs all characteristics of the first introductory course and instructor. The second-instructor variables remain identified because students from the same first-course section subsequently enroll in different second-course sections taught by different instructors.

The selection test on the gender-match treatment is one piece of a broader test. Appendix Figure A1 reports standardized differences in pre-treatment student characteristics across the two subsamples defined by the gender of the second-semester instructor, residualized within first-semester assignment cells in the same way as Figure 1. All but one covariate fall comfortably within the conventional negligible-imbalance range of $|d| < 0.10$, with most well below

0.06 in absolute value. The remaining covariate, transfer credits, has a standardized difference of 0.105, only marginally above that heuristic benchmark and well below the conventional substantial-imbalance threshold of 0.25. The balance pattern is therefore comparable to the first-semester assignment evidence in Figure 1.

To translate the measured selection on observables into a quantitative constraint on the second-instructor coefficient, we follow the coefficient-stability approach of Oster (2019), which builds on the proportional-selection framework of Altonji et al. (2005). Appendix Table A15 reports, for a grid of hypothesized true cumulative-exposure effects equal to a fraction α of the first-instructor effect, the bias on the second-instructor coefficient that would be required to drive the observed near-zero estimate. Under the standard proportionality assumption, masking a true cumulative-exposure effect equal to half the first-instructor effect on advanced-course enrollment requires the implied selection on unobservables to run roughly nine times stronger than the measured selection on observables, and in the opposite direction.[7] Under proportional selection, the implied degree of unobserved selection required to mask an economically meaningful cumulative-exposure effect is substantially larger than the already modest observed selection in the data.

The first-instructor coefficient is unmoved by the second-instructor block, the second-instructor block is jointly indistinguishable from zero, and the pattern holds across all three fixed-effects specifications and both clustering schemes. Under cumulative exposure with positive treatment effects, students who responded to a female first instructor would be expected to selectively sort into female-taught second sections, biasing the OLS coefficient on the second-instructor block upward and therefore making the test conservative against finding no cumulative-exposure effect. In the opposite scenario, where positive first exposure leads some treated women to avoid female second instructors, selection would be negative and the null could mask a positive cumulative effect. The Oster bound suggests that the magnitudes of unobserved selection required for that scenario to operate are implausibly large relative to the observed selection in the data. Cumulative effects could still matter for narrow subgroups not identified by this design, but the estimates provide little evidence that cumulative exposure operates at substantively important magnitudes in this setting.

The pattern is consistent with several mechanisms that the design cannot separately identify, including belief updating and identity-related responses to an early encounter with a female instructor (Akerlof and Kranton, 2000; Bordalo et al., 2019; Manski, 1993). Across these interpretations, the operative policy margin is initial section assignment, the implementable margin for departments moving female faculty toward parity in gateway courses. The estimates therefore speak more clearly to the importance of first exposure than to the precise psychological or informational mechanism producing it. The interpretation is nevertheless

[7] The proportionality concept of Oster (2019) and Altonji et al. (2005) expresses the required bias as a multiple δ of the movement that adding the pre-treatment observables block produces in the second-instructor coefficient (Appendix Table A15). The yardstick movement is -0.0054 for advanced-course enrollment, giving the implied $\delta = -8.75$ at $\alpha = 0.50$. The negative sign indicates that selection on unobservables would have to run in the opposite direction of the measured selection on observables, requiring treated women to avoid female second instructors despite responding positively to female first instructors. For major completion the yardstick is mechanically near zero, so the standard-error-based diagnostic (1.5 standard errors of the observed second-instructor estimate) is the more informative metric.

consistent with evidence that even relatively brief exposure to female teachers can alter girls' educational expectations and later-life trajectories in settings where female professional representation is limited (Card et al., 2022).

5.3.2 Grades and the Mediation Bound

Prior research suggests that gender disparities in undergraduate economics partly reflect women's greater responsiveness to grades, with women becoming less likely to continue after receiving lower-than-expected marks even at higher levels of measured performance (Rask and Tiefenthaler, 2008). Related work also finds that women often underestimate their quantitative ability (Emerson et al., 2012). A separate literature shows that teacher-assigned grades can reflect factors beyond test performance (Terrier, 2020; Cornwell et al., 2013; Elder and Zhou, 2021), leaving open whether instructor effects on grades reflect changes in student performance, changes in grading, or both.

We re-estimate equation 1 using introductory-course performance as the outcome (Table 7). We examine both GPA points earned in the course on the UW–La Crosse grading scale[8] and an indicator for earning an A or AB, which captures top-of-distribution performance. All specifications use the same controls, fixed effects, and clustering structure as the baseline estimates. The table shows that female instructors close part of the within-cell gender gap in course performance. In Panel A, female students taught by male instructors earn roughly 0.10 GPA points less than their male peers. Under female instructors, the female-student point estimates are modestly positive but imprecise, leaving the within-cell gender gap close to zero. The implied female-instructor effect on the gap is 0.15 to 0.16 GPA points, on the order of 15 percent of a full letter grade. Panel B, which uses an indicator for receiving an A or AB, reports a smaller and largely imprecise pattern. The within-cell GPA effect is therefore concentrated in the middle of the grade distribution rather than at the top.

The persistence effect cannot be point-identified into grade-mediated and grade-independent components without auxiliary assumptions, since grades are post-treatment (e.g., Angrist and Pischke, 2009; Imai et al., 2010). Rather than impose sequential ignorability of grades, we use a transparent accounting decomposition anchored to external estimates of the grade-to-persistence relationship from comparable introductory-economics settings. Let F denote the female student-female instructor interaction and H an indicator for earning a B or better in the first-semester introductory course. The reduced-form persistence effect,

$$\Delta Y = E[Y \mid F = 1] - E[Y \mid F = 0],$$

and the treatment effect on grades,

$$\Delta H = E[H \mid F = 1] - E[H \mid F = 0],$$

are identified from our data. Following the mediation framework of Imai et al. (2010), we decompose the reduced-form gender-match effect into a component operating through grades

[8] The grading scale assigns A = 4.0, AB = 3.5, B = 3.0, BC = 2.5, C = 2.0, D = 1.0, F = 0.0. The half-letter grades AB and BC are institution-specific and capture finer gradations than the standard A-through-F scale.

and a residual non-grade component:

$$\Delta Y = \underbrace{\Delta Y^{\text{direct}}}_{\text{non-grade component}} + \underbrace{\left(\frac{\partial Y}{\partial H}\right) \Delta H}_{\text{through grades}}.$$

In Table 8, we calibrate $\partial Y/\partial H$ using external estimates from introductory-economics settings and then infer the implied residual non-grade component.

The primary anchor is Antman et al. (2025), whose introductory-economics sequence at a U.S. public university most closely matches ours and whose relative-grade-signal information design identifies a population partial rather than a cutoff-local effect. At their modal estimate of $\partial Y/\partial H = 0.10$, the grade channel absorbs roughly 11 percent of the gender-match effect on advanced-course enrollment and 29 percent of the smaller effect on major completion. Across the broader literature-anchor grid $\partial Y/\partial H \in [0.025, 0.30]$, spanning RD estimates from Owen (2010), Main and Ost (2014), and McEwan et al. (2021), the implied indirect share on advanced-course enrollment lies in $[3, 33]$ percent and the residual non-grade component remains positive at every anchor. The qualitative reading holds row by row. These RD-derived anchors identify cutoff-local effects rather than population partials, so treating them as the average partial across the support of H is an approximation, and the upper-tail bounds reported here should be read as conservative under that approximation.[9]

A product-of-coefficients exercise under sequential ignorability (Imai et al., 2010) yields a mediated share of at most 15 percent on advanced-course enrollment and 23 percent on major completion, with sensitivity parameter $\rho^* \approx 0.05$ at which the mediated share falls to zero.[10] The implied mediation share is therefore highly sensitive to even modest unobserved confounding between grades and persistence.

We next examine whether the gender-match effect varies systematically with predicted academic preparation. Appendix D constructs a predicted-grade index $\hat{G}(X)$ using pre-treatment characteristics, including ACT score, math placement, demographic indicators, and transfer credits. The gender-match effect is essentially flat across quartiles of $\hat{G}(X)$ ($p = 0.80$ for advanced course-taking and $p = 0.55$ for major completion). By contrast, the relationship between grades and persistence differs across quartiles: among female students, the grade-to-persistence slope is near zero in the lowest predicted-grade quartile and positive in the upper three, with across-quartile equality rejected at conventional levels for both outcomes. A predominantly grade-driven explanation would therefore require the gender-match effect and the grade-to-persistence relationship to vary similarly across predicted preparation, which the data do not support.

[9]Anchors above 0.10 lean on regression-discontinuity estimates and treat a cutoff-local effect as the average partial across the support of H. If compliers near a grade discontinuity have larger $\partial Y/\partial H$ than the population average, the upper-tail anchors overstate the indirect share and the bounds reported here are conservative. The Antman et al. (2025) modal anchor is not subject to this concern. Appendix D reports the full grid, the female-only and male-only mirror specifications, and the underlying point-estimate tables.

[10]ρ^* is the correlation between unobserved determinants of the mediator and the outcome at which the average causal mediation effect equals zero. Standard errors via the delta method. The implication of $\rho^* \approx 0.05$ is that almost any unobserved determinant of the introductory-economics grade that also affects persistence (effort, interest in the field, classroom engagement, contemporaneous shocks) would suffice to eliminate the mediated share entirely.

These exercises rely on different identifying assumptions, but all point to the same qualitative conclusion. Grades account for only a limited share of the gender-match effect on advanced-course enrollment. The pattern is broadly consistent with role-model and identity-related interpretations, particularly in environments where women are underrepresented. The findings are difficult to reconcile with explanations based primarily on grades alone.

6 Conclusion

Who teaches the first introductory economics course matters for who continues in the field. In a setting where students are placed into sections separately from instructor assignments, female introductory instructors close roughly 40 percent of the gender gap in advanced-course enrollment. The estimate for major completion is smaller and noisier but points the same way. Male students show no response, which weighs against the idea that female instructors are simply better teachers.

The effect shows up in the first course rather than building with later exposure. Students who later take a section with a female instructor are not more likely to continue. Grades play only a small part. Female instructors raise women's grades, but most of the change in whether women continue operates through other channels. The pattern fits explanations that emphasize identity, belonging, and how students update their views about whether they belong in the field (Akerlof and Kranton, 2000; Card et al., 2022; Manski, 1993; Wiswall and Zafar, 2015). It is also consistent with survey evidence that women question whether they fit in economics (Buzard et al., 2025).

Several limitations bear noting. The analysis uses data from a single mid-sized public university, and although College Scorecard benchmarking places UWL alongside roughly 100 peer institutions on observable dimensions, the patterns may not extend to settings with very different student or faculty composition. The instructor pool is small (14 female and 14 male instructors), which limits power for subgroup analyses and leaves open whether the women teaching gateway courses at UWL are stylistically representative of female introductory-economics faculty more broadly. The design cannot separately identify which non-grade channel produces the persistence response, only that grades account for a limited share of it.

The findings point to a practical lever. Departments can shape the economics pipeline through who teaches the first course rather than through programs that reach only a subset of students. The binding constraint is the supply of female faculty. At departments where women teach only about a third of introductory sections (Lundberg and Stearns, 2019), a move to parity would increase exposure by 17 percentage points. Applied to a UWL-sized cohort of 109 women in introductory economics, this implies about 1.1 additional women per cohort taking an advanced course (a 7% increase) and 0.4 additional majors (a 17% increase). The calculation assumes that the female faculty teaching gateway courses at lower-representation departments are similar to those at near-parity departments.

References

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1):151–184.

American Association of State Colleges and Universities (2026). Issue summary: Regional public universities. https://aascu.org/resources/issue-summary-regional-public-universities/. Accessed May 27, 2026.

Andersen, I. G. and Reimer, D. (2019). Same-gender teacher assignment, instructional strategies, and student achievement: New evidence on the mechanisms generating same-gender teacher effects. *Research in Social Stratification and Mobility*, 62:100406.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

Antman, F. M., Skoy, E., and Flores, N. E. (2025). Can better information reduce college gender gaps? the impact of relative grade signals on academic outcomes for students in introductory economics. *Southern Economic Journal*.

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., and Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *Quarterly Journal of Economics*, 131(3):1415–1453.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28(25):3083–3107.

Bayer, A. and Rouse, C. E. (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives*, 30(4):221–242.

Bettinger, E. P. and Long, B. T. (2005). Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review Papers and Proceedings*, 95(2):152–157.

Board of Governors of the Federal Reserve System (2021). Changes in women's representation in economics: New data from the AEA papers and proceedings. FEDS Notes.

Bordalo, P., Coffman, K. B., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–773.

Buzard, K., Cranney, K., Gee, L. K., and Stoddard, O. (2025). What deters women from economics. *AEA Papers and Proceedings*, 115:659–663.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.

Canaan, S. and Mouganie, P. (2021). Does advisor gender affect women's persistence in economics? *AEA Papers and Proceedings*, 111:112–116.

Card, D., Giuliano, L., Løken, K. V., and Salvanes, K. G. (2022). Female teachers and girls' outcomes. NBER Working Paper 30430, National Bureau of Economic Research.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics*, 125(3):1101–1144.

Carrell, S. E. and West, J. E. (2010). Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3):409–432.

Chambers, A., Dickert-Conlin, S., Elder, C., Haider, S. J., and Imberman, S. (2021). Info.econ: Increasing diversity among economics majors. *AEA Papers and Proceedings*, 111:133–137.

Chevalier, J. (2021). Report: Committee on the status of women in the economics profession (cswep). *AEA Papers and Proceedings*, 111:742–763.

Cornwell, C. M., Mustard, D. B., and Parys, J. V. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1):236–264.

Elder, T. E. and Zhou, Y. (2021). The black-white gap in noncognitive skills among elementary school children. *American Economic Journal: Applied Economics*, 13(1):105–132.

Emerson, T. L. N., McGoldrick, K., and Mumford, K. J. (2012). Women and the choice to study economics. *Journal of Economic Education*, 43(4):349–362.

Good, P. I. (2006). *Resampling Methods: A Practical Guide to Data Analysis*. Birkhäuser, Boston, 3rd edition.

Hoffmann, F. and Oreopoulos, P. (2009). A professor like me: The influence of instructor gender on college achievement. *Journal of Human Resources*, 44(2):479–494.

Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51–71.

IPUMS USA (2025). Ipums usa: Version 15.0 [dataset].

Kofoed, M. S. and McGovney, E. (2019). The effect of same-gender or same-race role models on occupation choice: Evidence from randomly assigned mentors at west point. *Journal of Human Resources*, 54(2):430–467.

Lavy, V. (2016). What makes an effective teacher? quasi-experimental evidence. *CESifo Economic Studies*, 62(1):88–125.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Science & Business Media, 3rd edition.

Lim, J. and Meer, J. (2017). The impact of teacher–student gender matches: Random assignment evidence from south korea. *Journal of Human Resources*, 52(4):979–997.

Lim, J. and Meer, J. (2020). Persistent effects of teacher–student gender matches. *Journal of Human Resources*, 55(3):809–835.

Lundberg, S. and Stearns, J. (2019). Women in economics: Stalled progress. *AEA Papers and Proceedings*, 109:175–179.

MacKinnon, J. G. and Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254.

Main, J. B. and Ost, B. (2014). The impact of letter grades on student effort, course selection, and major choice: A regression-discontinuity analysis. *Journal of Economic Education*, 45(1):1–10.

Makles, A. (2012). Stata tip 110: How to get the optimal k-means cluster solution. *The stata journal*, 12(2):347–351.

Manski, C. F. (1993). Adolescent econometricians: How do youth infer the returns to schooling? In Clotfelter, C. T. and Rothschild, M., editors, *Studies of Supply and Demand in Higher Education*, pages 43–60. University of Chicago Press, Chicago.

Mansour, H., Rees, D., Rintala, B., and Wozny, N. (2022). The effects of professor gender on the post-graduation outcomes of female students. *Industrial and Labor Relations Review*, 75(3):693–715.

Maurer, S., Schwerdt, G., and Wiederhold, S. (2023). Do role models matter in large classes? new evidence on gender match effects in higher education. *CESifo Working Paper Series*, (10208).

May, A. M., McGarvey, M. G., Gustafson, C. R., and Mieno, T. (2021). Gender, environmental issues and policy: An examination of the views of male and female economists. *Ecological Economics*, 182:106877.

McEwan, P. J., Rogers, S., and Weerapana, A. (2021). Grade sensitivity and the economics major at a women's college. *AEA Papers and Proceedings*, 111:102–106.

National Center for Science and Engineering Statistics (2021). Women, minorities, and persons with disabilities in science and engineering: 2021. Technical Report NSF 21-321, National Science Foundation. Accessed May 28, 2026.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Owen, A. L. (2010). Grades, gender, and encouragement: A regression discontinuity analysis. *Journal of Economic Education*, 41(3):217–234.

Patnaik, A., Pauley, G. C., Venator, J., and Wiswall, M. J. (2024). The impacts of same and opposite gender alumni speakers on interest in economics. *Economics of Education Review*, 102:102579.

Porter, C. and Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, 12(3):226–254.

Pugatch, T. and Schroeder, E. (2021). Promoting female interest in economics: Limits to nudges. *AEA Papers and Proceedings*, 111:123–127.

Pugatch, T. and Schroeder, E. (2024). A simple nudge increases socioeconomic diversity in undergraduate economics. *Economic Inquiry*, 62(1):287–307.

Rask, K. and Tiefenthaler, J. (2008). The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review*, 27(6):676–687.

Roodman, D., MacKinnon, J. G., Nielsen, M. Ø., and Webb, M. D. (2019). Fast and wild: Bootstrap inference in stata using boottest. *The Stata Journal*, 19(1):4–60.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3):169–188.

Terrier, C. (2020). Boys lag behind: How teachers' gender biases affect student achievement. *Economics of Education Review*, 77:101981.

Westfall, P. H. and Young, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. John Wiley & Sons.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *Review of Economic Studies*, 82(2):791–824.

Xie, Y. and Shauman, K. A. (2004). Women in science: Career processes and outcomes. *Social Forces*, 82(4):1669–1671.

Tables

Table 1: Summary Statistics

	All Students		Male Students		Female Students	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Student Characteristics and Experiences						
White	0.9132	0.2816	0.9160	0.2774	0.9089	0.2879
First Generation	0.2942	0.4558	0.2728	0.4455	0.3269	0.4693
Pell Eligible	0.2207	0.4148	0.2195	0.4140	0.2224	0.4160
Attended Public HS	0.8787	0.3265	0.8628	0.3442	0.9030	0.2961
ACT Composite	24.4076	2.5889	24.5225	2.6188	24.2324	2.5337
Placement in Remedial Math	0.2021	0.4016	0.1959	0.3970	0.2115	0.4086
Placement in College Algebra	0.2273	0.4192	0.2195	0.4140	0.2391	0.4267
Placement in Pre-Calculus	0.3973	0.4894	0.4127	0.4925	0.3737	0.4840
Placement in Calculus I	0.1554	0.3623	0.1608	0.3675	0.1472	0.3544
Transfer Credits	5.3982	7.0757	4.5353	6.4254	6.7127	7.7860
First Semester Credits	15.0119	1.1491	14.9769	1.1521	15.0652	1.1429
College – Liberal Arts	0.0514	0.2208	0.0439	0.2049	0.0627	0.2425
College – Business	0.8526	0.3546	0.8557	0.3515	0.8478	0.3593
College – Sciences	0.0891	0.2850	0.0955	0.2940	0.0794	0.2705
Foreign Instructor	0.3595	0.4799	0.3491	0.4768	0.3754	0.4844
Instructor Experience (in years)	13.4629	11.7880	13.2525	11.7644	13.7834	11.8217
Tenured/Tenure-Track Instructor	0.5388	0.4986	0.5560	0.4970	0.5125	0.5001
Share of Females Enrolled in Class	0.3882	0.1014	0.3727	0.1014	0.4116	0.0968
Class Size	33.2916	7.8550	33.4259	7.9292	33.0870	7.7394
Panel B: Outcomes and Key Explanatory Variables						
Completed an Advanced Economics Course	0.2300	0.4209	0.2706	0.4444	0.1681	0.3741
Majored in Economics	0.0378	0.1907	0.0445	0.2062	0.0276	0.1639
Female	0.3963	0.4892	—	—	—	—
Female Instructor	0.5017	0.5001	0.4945	0.5001	0.5125	0.5001
N	3,018		1,822		1,196	

Notes: The table reports means and standard deviations for all students administratively pre-assigned to introductory microeconomics or macroeconomics between 2009–2019 at UWL. The unit of observation is the student; section- and instructor-level covariates are merged to the student via the student's assigned introductory section.

Table 2: University Comparisons

University Name	Number of Universities in Group	Admission Rate (%)	Undergraduate Population	Share White (%)	Share Female (%)	Net Price	Pell Receipt (%)	Instructional Expenditure per FTE	4-Year Degree Completed within 6 Years (%)	Student-to-Faculty Ratio	SAT Equivalent
Panel A: Our Setting											
University of Wisconsin-La Crosse (UWL)	1	0.7386	9,174	0.8929	0.5674	$14,633	0.2091	$6,495	0.6901	20	1,207
Panel B: Constructed Peer Groups for UWL											
University of Wisconsin–La Crosse Peer Group ($k = 4$)	186	0.7634	8,042	0.6964	0.5719	$12,867	0.3614	$7,137	0.4329	19	1,167
University of Wisconsin–La Crosse Peer Group ($k = 5$)	103	0.7188	11,050	0.7366	0.5406	$15,915	0.2927	$8,313	0.5809	19	1,209
University of Wisconsin–La Crosse Peer Group ($k = 6$)	81	0.6817	7,330	0.7410	0.5548	$16,770	0.2993	$8,767	0.5916	17	1,225
All Public Predominantly Undergraduate Universities	621	0.6818	9,888	0.5973	0.5545	$13,335	0.3633	$11,059	0.4906	17	1,165
Panel C: Other Universities Used in Role Model Studies											
Michigan State University (Chambers et al., 2021)	1	0.7019	37,649	0.7008	0.5081	$16,271	0.2279	$14,888	0.7812	17	1,124
Oregon State University (Pugatch and Schroeder, 2021)	1	0.7974	22,065	0.6793	0.4652	$17,492	0.2943	$10,144	0.6272	21	1,162
Southern Methodist University (Porter and Serra, 2020)	1	0.5161	6,294	0.6659	0.5073	—	0.1491	$15,826	0.7782	11	1,297
United States Air Force Academy (Carrell et al., 2010)	1	0.1360	4,268	0.6717	0.2308	—	0.0000	$27,039	0.8173	8	1,303
United States Military Academy (Kofoed and McGovney, 2019)	1	0.1060	4,535	0.6796	0.1797	—	0.0000	$37,638	0.8383	7	1,253
University of Wisconsin-Madison (Patnaik et al., 2024)	1	0.6017	29,288	0.7616	0.5132	$17,816	0.1437	$14,604	0.8386	19	1,297

Notes: The table reports summary statistics obtained from the U.S. Department of Education's *College Scorecard*. Values are presented for four groups: (i) the focal university that serves as the setting for this study; (ii) its peer institutions identified through clustering analyses based on institutional and student characteristics; (iii) a broader benchmark group of predominantly undergraduate public universities; and (iv) universities that have served as study settings in prior research on role models in education. Variables include measures of selectivity, student demographics, affordability, instructional expenditures, and academic outcomes. All monetary values are in U.S. dollars, and completion rates refer to the proportion of students earning a four-year degree within six years of initial enrollment.

Table 3: Gender-Match Effects in Introductory Economics Courses

	(1)	(2)	(3)
Panel A: Probability of Taking Advanced Economics Courses			
Gender Gap with Male Instructor	-0.1470	-0.1460	-0.1524
	$(0.0162)^{***}$	$(0.0170)^{***}$	$(0.0175)^{***}$
	$[0.0152]^{\dagger\dagger\dagger}$	$[0.0153]^{\dagger\dagger\dagger}$	$[0.0165]^{\dagger\dagger\dagger}$
Gender Gap with Female Instructor	-0.0818	-0.0841	-0.0840
	$(0.0231)^{***}$	$(0.0238)^{***}$	$(0.0242)^{***}$
	$[0.0287]^{\dagger\dagger\dagger}$	$[0.0275]^{\dagger\dagger\dagger}$	$[0.0291]^{\dagger\dagger\dagger}$
Difference in Gender Gaps	0.0652	0.0619	0.0684
	$(0.0277)^{**}$	$(0.0289)^{**}$	$(0.0288)^{**}$
	$[0.0327]^{\dagger\dagger}$	$[0.0308]^{\dagger\dagger}$	$[0.0316]^{\dagger\dagger}$
Panel B: Probability of Completing an Economics Major			
Gender Gap with Male Instructor	-0.0356	-0.0375	-0.0400
	$(0.0093)^{***}$	$(0.0093)^{***}$	$(0.0094)^{***}$
	$[0.0087]^{\dagger\dagger\dagger}$	$[0.0085]^{\dagger\dagger\dagger}$	$[0.0092]^{\dagger\dagger\dagger}$
Gender Gap with Female Instructor	-0.0143	-0.0133	-0.0133
	(0.0099)	(0.0098)	(0.0100)
	$[0.0101]$	$[0.0105]$	$[0.0106]$
Difference in Gender Gaps	0.0212	0.0241	0.0268
	(0.0131)	$(0.0131)^{*}$	$(0.0134)^{**}$
	$[0.0123]^{\dagger}$	$[0.0128]^{\dagger}$	$[0.0133]^{\dagger\dagger}$
Fixed Effects Included			
Term	Yes	No	No
Term×Course ×Days×Time	No	Yes	No
Class	No	No	Yes

Notes: The table reports percentage point differences in the probability that a student takes an advanced economics course (Panel A) or majors in economics (Panel B). Each specification includes 3,018 observations. Columns differ by the fixed effects listed at the bottom of the table. One-way clustered standard errors (by assignment cell) are shown in parentheses; two-way clustered standard errors (by assignment cell and instructor) are shown in brackets. All specifications include the full set of student, instructor, and class controls. Instructor and class controls, other than instructor experience, are absorbed by assignment-cell fixed effects or fully accounted for by class fixed effects. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

Table 4: Does Gender-Match Treatment Predict Choice of Second-Semester Instructor?

	Dependent Variable: Female Second-Semester Instructor		
	Term FE (1)	Assignment-Cell FE (2)	Class FE (3)
Female Student × Female Instructor (1st)	-0.0434	-0.0246	-0.0500
	(0.0389)	(0.0348)	(0.0351)
	[0.0316]	[0.0308]	[0.0313]
N	2,508	2,508	2,508

Notes: The dependent variable is an indicator equal to one if the second-semester introductory economics instructor is female. The sample is the same set of students used in the second-instructor robustness tables, restricted to those who completed both introductory courses and to observations retained under the most restrictive (class) fixed-effect specification, so all three columns use an identical set of observations. The key regressor is the gender-match interaction (Female Student × Female Instructor) from the first-semester course. Under the null of no selection, students who benefited from a female first-semester instructor should not differentially sort into female second-semester instructors. All specifications include the full set of student, instructor, and class controls. Standard errors clustered by assignment cell in parentheses, and two-way clustered by assignment cell and instructor in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Second-Instructor Robustness: Probability of Taking Advanced Economics Courses

	(1)	(2)	(3)
Comparison row: baseline gender-match coefficient on this subsample			
Female Student × Female Instructor [1st]	0.0759	0.0768	0.0870
	(0.0333)**	(0.0348)**	(0.0358)**
	[0.0411]†	[0.0398]†	[0.0394]††
Augmented specification (adds second-instructor controls)			
Female Student	−0.1577	−0.1626	−0.1666
	(0.0242)***	(0.0252)***	(0.0269)***
	[0.0275]†††	[0.0291]†††	[0.0282]†††
Female Instructor [1st]	−0.0257	0.0097	—
	(0.0259)	(0.0331)	—
	[0.0197]	[0.0310]	—
Female Student × Female Instructor [1st]	0.0748	0.0749	0.0857
	(0.0334)**	(0.0349)**	(0.0358)**
	[0.0397]†	[0.0380]††	[0.0388]††
Female Instructor [2nd]	−0.0155	−0.0163	−0.0186
	(0.0277)	(0.0286)	(0.0309)
	[0.0246]	[0.0280]	[0.0278]
Female Student × Female Instructor [2nd]	0.0042	0.0112	0.0031
	(0.0314)	(0.0313)	(0.0335)
	[0.0312]	[0.0353]	[0.0362]
F-test, second-instructor block = 0 (p-value, 1-way cluster)	0.8248	0.8497	0.7899
F-test, second-instructor block = 0 (p-value, 2-way cluster)	0.7585	0.8451	0.7249
N	2,508	2,508	2,508
Fixed Effects Included			
Term	Yes	No	No
Term×Course ×Days×Time	No	Yes	No
Class	No	No	Yes

Notes: The table tests whether advanced economics course-taking reflects first exposure to a female introductory instructor or cumulative dosage of female-instructor contact across the introductory sequence. The sample is restricted to students who took both introductory courses, approximately 83 percent of the analytic sample (2,508 of 3,018 master-sample observations). To allow direct comparison across columns, all three specifications use the same set of observations, defined by the intersection of the master sample, the both-courses subsample, and the observations retained under class fixed effects. *Female Instructor [1st]* is the gender of the first-semester (administratively assigned) introductory-economics instructor and corresponds to the gender-match treatment in our baseline specification. *Female Instructor [2nd]* is the gender of the other introductory-economics instructor, typically taken in a subsequent semester. The comparison row reports the baseline gender-match coefficient on the same both-courses subsample without the second-instructor controls. The augmented sub-panel adds the main effect and female-student interaction for the second instructor. Under a cumulative-exposure channel the second-instructor terms should attenuate the first-instructor coefficient relative to the comparison row, and under a first-exposure channel they should not. All specifications include the full set of student, instructor, and class controls. One-way clustered standard errors (by assignment cell) are shown in parentheses, and two-way clustered standard errors (by assignment cell and instructor) are shown in brackets. Class fixed effects (column 3) absorb all first-section identifiers, and the second-instructor variables remain identified within first-course classes because students from the same first-course section take their second introductory-economics course in different sections with potentially different instructors. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

Table 6: Second-Instructor Robustness: Probability of Completing an Economics Major

	(1)	(2)	(3)
Comparison row: baseline gender-match coefficient on this subsample			
Female Student × Female Instructor [1st]	0.0232	0.0297	0.0315
	(0.0163)	(0.0162)*	(0.0162)*
	[0.0156]	[0.0158]†	[0.0163]†
Augmented specification (adds second-instructor controls)			
Female Student	−0.0358	−0.0373	−0.0423
	(0.0135)***	(0.0140)***	(0.0144)***
	[0.0118]†††	[0.0122]†††	[0.0138]†††
Female Instructor [1st]	−0.0096	−0.0049	—
	(0.0133)	(0.0172)	—
	[0.0101]	[0.0156]	—
Female Student × Female Instructor [1st]	0.0236	0.0306	0.0314
	(0.0167)	(0.0166)*	(0.0165)*
	[0.0163]	[0.0162]†	[0.0169]†
Female Instructor [2nd]	0.0034	0.0061	0.0007
	(0.0115)	(0.0117)	(0.0129)
	[0.0116]	[0.0122]	[0.0141]
Female Student × Female Instructor [2nd]	−0.0019	−0.0057	0.0006
	(0.0163)	(0.0165)	(0.0164)
	[0.0142]	[0.0150]	[0.0149]
F-test, second-instructor block = 0 (p-value, 1-way cluster)	0.9542	0.8713	0.9953
F-test, second-instructor block = 0 (p-value, 2-way cluster)	0.9574	0.8406	0.9974
N	2,508	2,508	2,508
Fixed Effects Included			
Term	Yes	No	No
Term×Course ×Days×Time	No	Yes	No
Class	No	No	Yes

Notes: The table tests whether economics-major completion reflects first exposure to a female introductory instructor or cumulative dosage of female-instructor contact across the introductory sequence. The sample is restricted to students who took both introductory courses, approximately 83 percent of the analytic sample (2,508 of 3,018 master-sample observations). To allow direct comparison across columns, all three specifications use the same set of observations, defined by the intersection of the master sample, the both-courses subsample, and the observations retained under class fixed effects. *Female Instructor [1st]* is the gender of the first-semester (administratively assigned) introductory-economics instructor and corresponds to the gender-match treatment in our baseline specification. *Female Instructor [2nd]* is the gender of the other introductory-economics instructor, typically taken in a subsequent semester. The comparison row reports the baseline gender-match coefficient on the same both-courses subsample without the second-instructor controls. The augmented sub-panel adds the main effect and female-student interaction for the second instructor. Under a cumulative-exposure channel the second-instructor terms should attenuate the first-instructor coefficient relative to the comparison row, and under a first-exposure channel they should not. All specifications include the full set of student, instructor, and class controls. One-way clustered standard errors (by assignment cell) are shown in parentheses, and two-way clustered standard errors (by assignment cell and instructor) are shown in brackets. Class fixed effects (column 3) absorb all first-section identifiers, and the second-instructor variables remain identified within first-course classes because students from the same first-course section take their second introductory-economics course in different sections with potentially different instructors. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

Table 7: Gender-Match Effects on Course Performance

	(1)	(2)	(3)
Panel A: Course GPA (Points)			
Gender Gap with Male Instructor	-0.0996	-0.1103	-0.0983
	$(0.0414)^{**}$	$(0.0380)^{***}$	$(0.0382)^{**}$
	$[0.0391]^{\dagger\dagger}$	$[0.0373]^{\dagger\dagger\dagger}$	$[0.0425]^{\dagger\dagger}$
Gender Gap with Female Instructor	0.0544	0.0540	0.0626
	(0.0447)	(0.0455)	(0.0470)
	[0.0484]	[0.0546]	[0.0510]
Difference in Gender Gaps	0.1540	0.1643	0.1609
	$(0.0606)^{**}$	$(0.0583)^{***}$	$(0.0596)^{***}$
	$[0.0613]^{\dagger\dagger}$	$[0.0650]^{\dagger\dagger}$	$[0.0642]^{\dagger\dagger}$
Panel B: Probability of Receiving A or AB			
Gender Gap with Male Instructor	-0.0169	-0.0200	-0.0243
	(0.0226)	(0.0217)	(0.0208)
	[0.0186]	[0.0205]	[0.0200]
Gender Gap with Female Instructor	0.0089	0.0121	0.0231
	(0.0224)	(0.0220)	(0.0223)
	[0.0206]	[0.0221]	[0.0197]
Difference in Gender Gaps	0.0257	0.0321	0.0473
	(0.0309)	(0.0298)	(0.0288)
	[0.0263]	[0.0285]	$[0.0258]^{\dagger}$
Fixed Effects Included			
Term	Yes	No	No
Term×Course ×Days×Time	No	Yes	No
Class	No	No	Yes

Notes: The table reports estimated effects of student-instructor gender match on course performance. The Panel A outcome is the student's GPA points in the introductory economics course on the UW-La Crosse grading scale (A = 4.0, AB = 3.5, B = 3.0, BC = 2.5, C = 2.0, D = 1.0, F = 0.0). The Panel B outcome is an indicator equal to one if the student received a course grade of A or AB (gpa $\geq$ 3.5). Each specification uses the same student, instructor, and class controls and the same fixed-effect and clustering structure as the baseline gender-match estimates in Table 3. The sample includes 3,018 student-section observations matched to the baseline analytic sample. One-way clustered standard errors (by assignment cell) are shown in parentheses, and two-way clustered standard errors (by assignment cell and instructor) are shown in brackets. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

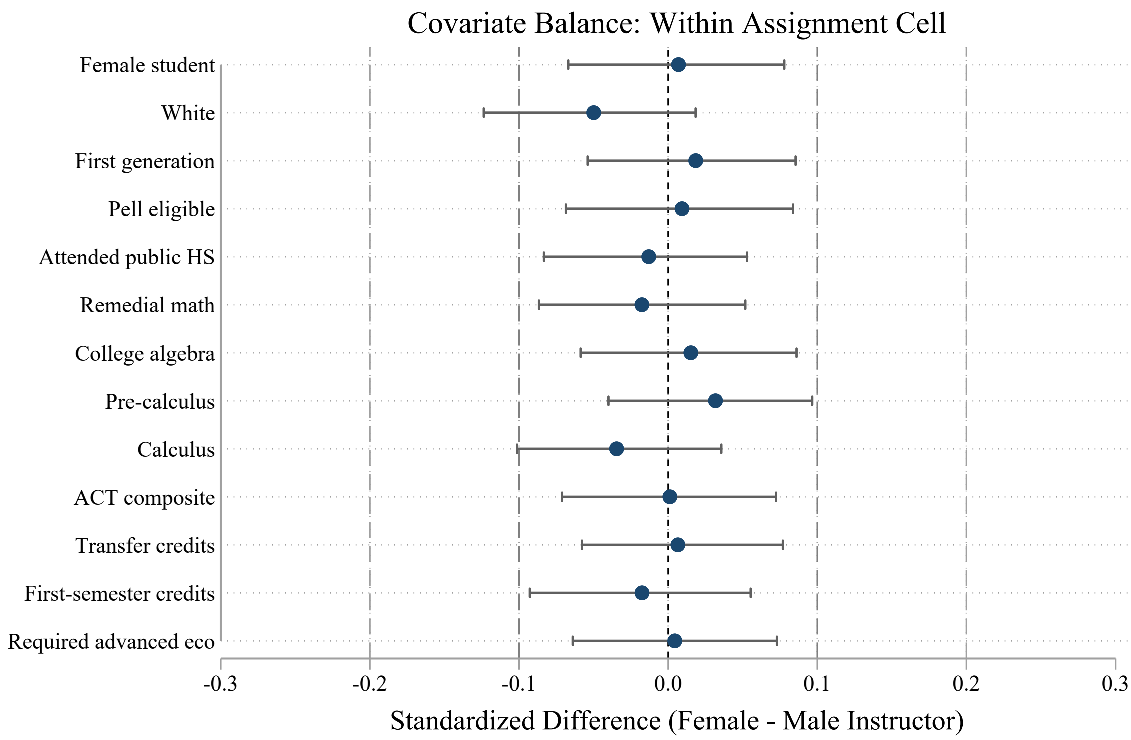
Table 8: Literature-Anchor Bounds on the Grade Channel

	Reduced Form (pp)		Implied Non-Grade Component (pp) [Indirect Share]			
	ΔY	ΔH	$\partial Y/\partial H = 0.05$	0.10	0.20	0.30
Advanced Course	6.2	6.9	5.8 [6%]	5.5 [11%]	4.8 [22%]	4.1 [33%]
Major Completion	2.4	6.9	2.1 [14%]	1.7 [29%]	1.0 [57%]	0.3 [86%]

Notes. The decomposition $\Delta Y = \Delta Y^{\text{direct}} + (\partial Y/\partial H) \cdot \Delta H$ is applied to the pooled-sample assignment-cell interaction specification ($N = 3{,}018$). The mediator H is an indicator for earning a B or higher in the introductory course. ΔY is the gender-match interaction effect on the persistence outcome, and ΔH is the corresponding effect on the mediator. Each cell in the right block reports the implied non-grade component $\Delta Y^{\text{direct}} = \Delta Y - (\partial Y/\partial H) \cdot \Delta H$ at the column anchor $\partial Y/\partial H$, with the indirect share $(\partial Y/\partial H) \cdot \Delta H/\Delta Y$ in brackets. The 0.10 modal anchor follows Antman et al. (2025), whose introductory-economics setting and information-experiment design most closely match ours. The 0.20 and 0.30 anchors draw on Owen (2010), Main and Ost (2014), and McEwan et al. (2021), which use regression-discontinuity designs and identify a local effect at the grade cutoff rather than the average partial across the support of H. The upper-tail anchors should be read as approximations under the additional assumption that the local-RD effect is informative for the population partial. The exercise reports bounds rather than point identification, under the assumption that grade is the only mediator and that the additivity of the decomposition holds. Appendix D reports the full anchor grid, point-estimate tables, and the female-only and male-only mirror specifications.

Figures

Figure 1: Standardized Imbalance — Student Assignment to Male versus Female Instructors



Notes: The figure plots standardized mean differences in student characteristics between those assigned to female versus male instructors in their first semester of introductory economics. Standardized differences are computed within assignment cells (term $\times$ course $\times$ day $\times$ time): each covariate is residualized on its cell mean, and the standardized difference is the residualized group-mean difference (Female - Male) divided by the pooled standard deviation. Horizontal bars are 95% bootstrap confidence intervals based on 1,000 resamples (seed 20260415). Dashed vertical lines at ± 0.10 and ± 0.20 SD mark conventional thresholds for negligible and moderate imbalance, and the solid line at zero indicates perfect balance. All covariates fall within ± 0.10, indicating that, conditional on assignment cell, students assigned to female and male instructors are observably similar across demographic, academic preparation, and enrollment characteristics. Sample sizes are approximately 1,500 students per group, drawn from the 3,018-observation analytic sample.

Appendix A — Supplementary Tables and Estimates

Table A1: Variable Definitions

Variable	**Definition**
Panel A: Student Characteristics and Experiences	
White	=1 if race/ethnicity reported as White (non-Hispanic) in institutional records.
First Generation	=1 if neither parent holds a four-year college degree.
Pell Eligible	=1 if the student was eligible for or received a Federal Pell Grant in the first year.
Attended Public HS	=1 if last high school attended was public.
ACT Composite	ACT composite score (highest on record prior to matriculation).
Math Placements	Mutually exclusive indicators based on the university's math placement rules
Transfer Credits	Number of credits accepted at entry (AP/IB/dual-enrollment/other transfer).
First Semester Credits	Total credit load attempted in the student's first term.
Required Advanced Economics	=1 if the student's declared major requires at least one advanced economics course; 0 otherwise.
Application-Date Decile	Decile of the student's undergraduate application date within their entering term. Included as factor variables in the regressions to absorb timing effects.
Registration-Date Decile	Decile of the student's class-registration date within their entering term. Included as factor variables.
College	Mutually exclusive indicators for the student's home college (Business, Liberal Arts, Sciences, Other) at first enrollment.
Foreign Instructor	=1 if the assigned instructor is foreign-born according to HR/personnel records.
Instructor Experience (years)	Years of post-secondary teaching experience recorded for the assigned instructor as of the term taught.
Tenured/Tenure-Track Instructor	=1 if the assigned instructor held a tenured or tenure-track appointment in that term.
Share of Females Enrolled in Class	Fraction of women in the student's assigned section.
Class Size	Total number of students enrolled in the assigned section.
Panel B: Outcomes and Key Explanatory Variables	
Completed an Advanced Economics Course	=1 if the student enrolled in any economics course beyond the introductory principles course(s) at any time after the first exposure; 0 otherwise.
Majored in Economics	=1 if an economics major was recorded at graduation; 0 otherwise.
Female Student	=1 if student is a female; 0 otherwise.
Female Instructor	=1 if the student's introductory section was taught by a female instructor; 0 otherwise.

Table A2: Covariate Balance on Instructor Gender

	Standardized Difference		Coefficient on Female Instr.	
Covariate	Raw	Within-cell	Coef.	p-value
Panel A: Student characteristics				
Female student	0.035	0.007	0.0217	0.389
White	-0.140	-0.050	-0.0342	0.368
First generation	-0.039	0.019	0.0210	0.696
Pell eligible	0.016	0.009	0.0054	0.864
Attended public HS	-0.034	-0.013	-0.0043	0.772
Remedial math	-0.061	-0.018	-0.0069	0.320
College algebra	-0.007	0.015	0.0148	0.609
Pre-calculus	0.095	0.032	0.0352	0.220
Calculus	-0.058	-0.035	-0.0431	0.275
ACT composite	0.014	0.001	0.0439	0.776
Transfer credits	0.042	0.007	0.2410	0.524
First-semester credits	0.001	-0.018	-0.0382	0.627
Required advanced eco	0.004	0.004	0.0007	0.921
Panel B: Instructor characteristics				
Foreign-born instructor	–	–	0.1887	0.021
Tenured / tenure-track	–	–	-0.2998	0.000
Instructor experience (yrs)	–	–	-0.1117	0.948
Section share female	–	–	0.0068	0.575
Class size	–	–	2.9854	0.049
Joint F-test (student covariates)			1.276	0.244
Section-level joint F-test			1.151	0.324

Notes: Panel A reports balance tests for student covariates. Panel B reports instructor attributes regressed on the female-instructor indicator. Raw standardized differences are computed on pooled data following Rosenbaum and Rubin (1985) and Austin (2009). Within-cell standardized differences are computed after partialling out assignment-cell fixed effects (Term $\times$ Course $\times$ Day $\times$ Time). The regression columns report the coefficient on the female-instructor indicator from a regression of each covariate on the female-instructor indicator, the assignment-cell fixed effects, and the timing controls, with standard errors clustered at the assignment cell. The joint F-test regresses the female-instructor indicator on all Panel A covariates. The section-level joint F-test aggregates to the assignment-cell mean. Imbalance thresholds are $|d| < 0.10$ negligible, $0.10 \leq |d| < 0.25$ small, and $|d| \geq 0.25$ substantial. The estimation sample is the analytic sample of $N = 3{,}018$ student-section observations.

Table A3: Sensitivity Checks and Regressions Used to Produce Estimates in Tables 3 and A12 (Dependent Variable: Probability of Taking an Advanced Economics Course)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Student	-0.1460^{***}	-0.1491^{***}	-0.1424^{***}	-0.1622^{***}	-0.1443^{***}	-0.1673^{***}	-0.1280^{***}	-0.1451^{***}
	(0.0170)	(0.0171)	(0.0172)	(0.0312)	(0.0241)	(0.0300)	(0.0262)	(0.0261)
Female Instructor	0.0179	—	—	−0.0264	0.0334	−0.0653	0.0057	−0.0206
	(0.0289)	—	—	(0.0403)	(0.0335)	(0.0548)	(0.0302)	(0.0457)
Female Student×	0.0619^{**}	0.0649^{**}	0.0533^{*}	0.1093^{**}	0.0400	0.0426	0.0864^{**}	0.0403
Female Instructor	(0.0289)	(0.0285)	(0.0294)	(0.0495)	(0.0344)	(0.0461)	(0.0396)	(0.0378)
Non-First Generation	—	—	—	−0.0234	—	—	—	—
	—	—	—	(0.0311)	—	—	—	—
Female Student×	—	—	—	0.0218	—	—	—	—
Non-First Generation	—	—	—	(0.0374)	—	—	—	—
Female Instructor×	—	—	—	0.0625	—	—	—	—
Non-First Generation	—	—	—	(0.0404)	—	—	—	—
Female Student ×Female Instructor	—	—	—	−0.0652	—	—	—	—
×Non-First Generation	—	—	—	(0.0541)	—	—	—	—
High Math Placement	—	—	—	—	0.0879	—	—	—
	—	—	—	—	(0.0580)	—	—	—
Female Student×	—	—	—	—	−0.0022	—	—	—
High Math Placement	—	—	—	—	(0.0351)	—	—	—
Female Instructor×	—	—	—	—	−0.0275	—	—	—
High Math Placement	—	—	—	—	(0.0368)	—	—	—
Female Student×Female Instructor	—	—	—	—	0.0399	—	—	—
×High Math Placement	—	—	—	—	(0.0510)	—	—	—
Foreign Instructor	—	—	—	—	—	-0.0681^{*}	—	—
	—	—	—	—	—	(0.0397)	—	—
Female Student×	—	—	—	—	—	0.0299	—	—
Foreign Instructor	—	—	—	—	—	(0.0375)	—	—
Female Instructor×	—	—	—	—	—	0.1004	—	—
Foreign Instructor	—	—	—	—	—	(0.0661)	—	—
Female Student×Female Instructor	—	—	—	—	—	0.0400	—	—
×Foreign Instructor	—	—	—	—	—	(0.0599)	—	—
High Female Share	—	—	—	—	—	—	−0.0131	—
	—	—	—	—	—	—	(0.0355)	—
Female Student×	—	—	—	—	—	—	−0.0341	—
High Female Share	—	—	—	—	—	—	(0.0436)	—
Female Instructor×	—	—	—	—	—	—	0.0255	—
High Female Share	—	—	—	—	—	—	(0.0419)	—
Female Student×Female Instructor	—	—	—	—	—	—	−0.0426	—
×High Female Share	—	—	—	—	—	—	(0.0611)	—
Intro. to Micro	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
Female Student×	—	—	—	—	—	—	—	−0.0018
Intro. to Micro	—	—	—	—	—	—	—	(0.0336)
Female Instructor×	—	—	—	—	—	—	—	0.0608
Intro. to Micro	—	—	—	—	—	—	—	(0.0651)
Female Student×Female Instructor	—	—	—	—	—	—	—	0.0353
×Intro. to Micro	—	—	—	—	—	—	—	(0.0563)
Fixed Effects Included:								
Assignment cell	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Instructor	No	Yes	No	No	No	No	No	No
Instructor by Term	No	No	Yes	No	No	No	No	No

Notes: The table reports percentage point differences in the probability that a student takes an advanced economics course. Each specification includes 3,018 observations. Columns differ by the fixed effects listed at the bottom of the table. One-way clustered standard errors (by assignment cell) are shown in parentheses. All specifications include the full set of student, instructor, and class controls. Instructor and class controls, other than instructor experience, are absorbed by assignment-cell fixed effects. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering.

Table A4: Sensitivity Checks and Regressions Used to Produce Estimates in Tables 3 and A12 (Dependent Variable: Probability of Completing an Economics Major)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Student	−0.0375***	−0.0379***	−0.0358***	−0.0497***	−0.0414***	−0.0311**	−0.0309***	−0.0305***
	(0.0093)	(0.0094)	(0.0099)	(0.0165)	(0.0133)	(0.0151)	(0.0111)	(0.0112)
Female Instructor	−0.0011	—	—	0.0003	−0.0019	−0.0218	0.0010	−0.0200
	(0.0138)	—	—	(0.0214)	(0.0191)	(0.0208)	(0.0140)	(0.0180)
Female Student×	0.0241*	0.0254*	0.0205	0.0478*	0.0155	0.0099	0.0298*	0.0279
Female Instructor	(0.0131)	(0.0133)	(0.0138)	(0.0270)	(0.0188)	(0.0193)	(0.0167)	(0.0212)
Non-First Generation	—	—	—	−0.0114	—	—	—	—
	—	—	—	(0.0187)	—	—	—	—
Female Student×	—	—	—	0.0188	—	—	—	—
Non-First Generation	—	—	—	(0.0209)	—	—	—	—
Female Instructor×	—	—	—	−0.0015	—	—	—	—
Non-First Generation	—	—	—	(0.0255)	—	—	—	—
Female Student ×Female Instructor	—	—	—	−0.0352	—	—	—	—
×Non-First Generation	—	—	—	(0.0345)	—	—	—	—
High Math Placement	—	—	—	—	−0.0055	—	—	—
	—	—	—	—	(0.0168)	—	—	—
Female Student×	—	—	—	—	0.0066	—	—	—
High Math Placement	—	—	—	—	(0.0161)	—	—	—
Female Instructor×	—	—	—	—	0.0011	—	—	—
High Math Placement	—	—	—	—	(0.0194)	—	—	—
Female Student×Female Instructor	—	—	—	—	0.0170	—	—	—
×High Math Placement	—	—	—	—	(0.0262)	—	—	—
Foreign Instructor	—	—	—	—	—	−0.0257	—	—
	—	—	—	—	—	(0.0179)	—	—
Female Student×	—	—	—	—	—	−0.0090	—	—
Foreign Instructor	—	—	—	—	—	(0.0180)	—	—
Female Instructor×	—	—	—	—	—	0.0250	—	—
Foreign Instructor	—	—	—	—	—	(0.0227)	—	—
Female Student×Female Instructor	—	—	—	—	—	0.0225	—	—
×Foreign Instructor	—	—	—	—	—	(0.0272)	—	—
High Female Share	—	—	—	—	—	—	0.0112	—
	—	—	—	—	—	—	(0.0207)	—
Female Student×	—	—	—	—	—	—	−0.0139	—
High Female Share	—	—	—	—	—	—	(0.0200)	—
Female Instructor×	—	—	—	—	—	—	−0.0072	—
High Female Share	—	—	—	—	—	—	(0.0230)	—
Female Student×Female Instructor	—	—	—	—	—	—	−0.0076	—
×High Female Share	—	—	—	—	—	—	(0.0245)	—
Intro. to Micro	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
Female Student×	—	—	—	—	—	—	—	−0.0116
Intro. to Micro	—	—	—	—	—	—	—	(0.0174)
Female Instructor×	—	—	—	—	—	—	—	0.0311
Intro. to Micro	—	—	—	—	—	—	—	(0.0256)
Female Student×Female Instructor	—	—	—	—	—	—	—	−0.0059
×Intro. to Micro	—	—	—	—	—	—	—	(0.0278)
Fixed Effects Included:								
Assignment cell	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Instructor	No	Yes	No	No	No	No	No	No
Instructor by Term	No	No	Yes	No	No	No	No	No

Notes: The table reports percentage point differences in the probability that a student majors in economics. Each specification includes 3,018 observations. Columns differ by the fixed effects listed at the bottom of the table. One-way clustered standard errors (by assignment cell) are shown in parentheses. All specifications include the full set of student, instructor, and class controls. Instructor and class controls, other than instructor experience, are absorbed by assignment-cell fixed effects. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering.

Table A5: Robustness: Adding Female-Student $\times$ Instructor- and Class-Attribute Interactions (Dependent Variable: Probability of Taking an Advanced Economics Course)

	(1)	(2)	(3)
Gender Gap with Male Instructor	-0.1474^{***}	-0.1463^{***}	-0.1530^{***}
	$(0.0185)^{\dagger\dagger\dagger}$	$(0.0189)^{\dagger\dagger\dagger}$	$(0.0183)^{\dagger\dagger\dagger}$
Gender Gap with Female Instructor	-0.0782^{***}	-0.0800^{***}	-0.0791^{***}
	$(0.0233)^{\dagger\dagger\dagger}$	$(0.0231)^{\dagger\dagger\dagger}$	$(0.0245)^{\dagger\dagger\dagger}$
Difference in Gender Gaps	0.0692^{**}	0.0664^{**}	0.0739^{**}
	$(0.0305)^{\dagger\dagger}$	$(0.0302)^{\dagger\dagger}$	$(0.0306)^{\dagger\dagger}$
Auxiliary Interactions Added			
Female Student $\times$ Foreign Instructor	-0.0371	-0.0334	-0.0419
	(0.0302)	(0.0301)	(0.0310)
Female Student $\times$ Tenured	-0.0332	-0.0274	-0.0267
	(0.0227)	(0.0220)	(0.0236)
Female Student $\times$ Instructor Experience	-0.0025^{*}	-0.0025^{*}	-0.0023
	$(0.0010)^{\dagger\dagger\dagger}$	$(0.0009)^{\dagger\dagger\dagger}$	$(0.0010)^{\dagger\dagger}$
Female Student $\times$ Class Size	-0.0004	-0.0002	0.0001
	(0.0012)	(0.0014)	(0.0016)
Female Student $\times$ Share Female	-0.1829	-0.2287	-0.2624
	(0.1746)	(0.1892)	(0.1782)
Fixed Effects Included			
Term	Yes	No	No
Term$\times$Course $\times$Days$\times$Time	No	Yes	No
Class	No	No	Yes

Notes: The table reports robustness of the gender-match estimates from Table 3 when interactions of the female-student indicator with each instructor- and class-level attribute are added to the regression. Each specification includes 3,018 observations. Columns differ by the fixed effects listed at the bottom of the table. The Gender Gap with Male / Female Instructor rows are evaluated at sample means of the attributes via the delta method, and the Difference in Gender Gaps row is the coefficient on the Female Student $\times$ Female Instructor interaction in the augmented regression. One-way clustered standard errors (by assignment cell) determine the significance stars on the coefficient estimates. Standard errors in parentheses are two-way clustered by assignment cell and instructor. *, **, *** denote significance at 10, 5, 1% under one-way clustering, and †, ††, ††† denote the same under two-way clustering. All specifications include the full set of student, instructor, and class controls. Instructor and class controls other than instructor experience are absorbed by assignment-cell fixed effects or fully accounted for by class fixed effects.

Table A6: Robustness: Adding Female-Student $\times$ Instructor- and Class-Attribute Interactions (Dependent Variable: Probability of Completing an Economics Major)

	(1)	(2)	(3)
Gender Gap with Male Instructor	-0.0319^{***}	-0.0343^{***}	-0.0373^{***}
	$(0.0082)^{\dagger\dagger\dagger}$	$(0.0081)^{\dagger\dagger\dagger}$	$(0.0078)^{\dagger\dagger\dagger}$
Gender Gap with Female Instructor	-0.0165	-0.0147	-0.0135
	(0.0110)	(0.0115)	(0.0112)
Difference in Gender Gaps	0.0154	0.0196	0.0238^{*}
	(0.0129)	(0.0134)	$(0.0132)^{\dagger}$
Auxiliary Interactions Added			
Female Student $\times$ Foreign Instructor	0.0081	0.0056	0.0031
	(0.0126)	(0.0129)	(0.0125)
Female Student $\times$ Tenured	-0.0066	-0.0024	-0.0005
	(0.0077)	(0.0078)	(0.0086)
Female Student $\times$ Instructor Experience	0.0001	0.0002	-0.0001
	(0.0005)	(0.0004)	(0.0004)
Female Student $\times$ Class Size	0.0016^{***}	0.0017^{***}	0.0017^{***}
	$(0.0005)^{\dagger\dagger\dagger}$	$(0.0005)^{\dagger\dagger\dagger}$	$(0.0004)^{\dagger\dagger\dagger}$
Female Student $\times$ Share Female	-0.0877	-0.1072	-0.1488^{**}
	(0.0700)	(0.0706)	$(0.0665)^{\dagger\dagger}$
Fixed Effects Included			
Term	Yes	No	No
Term$\times$Course $\times$Days$\times$Time	No	Yes	No
Class	No	No	Yes

Notes: The table reports robustness of the gender-match estimates from Table 3 when interactions of the female-student indicator with each instructor- and class-level attribute are added to the regression. Each specification includes 3,018 observations. Columns differ by the fixed effects listed at the bottom of the table. The Gender Gap with Male / Female Instructor rows are evaluated at sample means of the attributes via the delta method, and the Difference in Gender Gaps row is the coefficient on the Female Student $\times$ Female Instructor interaction in the augmented regression. One-way clustered standard errors (by assignment cell) determine the significance stars on the coefficient estimates. Standard errors in parentheses are two-way clustered by assignment cell and instructor. *, **, *** denote significance at 10, 5, 1% under one-way clustering, and †, ††, ††† denote the same under two-way clustering. All specifications include the full set of student, instructor, and class controls. Instructor and class controls other than instructor experience are absorbed by assignment-cell fixed effects or fully accounted for by class fixed effects.

Table A7: Robustness: Adding Female Student × Baseline-Covariate and Female Student × Term Interactions

	(1)	(2)	(3)
Panel A: Probability of Taking Advanced Economics Courses			
Comparison: baseline gender-match coefficient			
Female Student × Female Instructor	0.0652	0.0619	0.0684
	(0.0277)**	(0.0289)**	(0.0288)**
	[0.0327]††	[0.0308]††	[0.0316]††
Augmented: adds female-student × covariate and × term interactions			
Female Student × Female Instructor	0.0846	0.0818	0.0919
	(0.0310)***	(0.0321)**	(0.0317)***
	[0.0354]††	[0.0341]††	[0.0350]†††
F-test, female-student × covariate block = 0 (p-value, 1-way cluster)	0.000	0.000	0.000
F-test, female-student × covariate block = 0 (p-value, 2-way cluster)	0.000	0.000	0.000
F-test, female-student × term block = 0 (p-value, 1-way cluster)	0.085	0.063	0.032
F-test, female-student × term block = 0 (p-value, 2-way cluster)	0.193	0.191	0.014
N	3,018	3,018	3,017
Panel B: Probability of Completing an Economics Major			
Comparison: baseline gender-match coefficient			
Female Student × Female Instructor	0.0212	0.0241	0.0268
	(0.0131)	(0.0131)*	(0.0134)**
	[0.0123]†	[0.0128]†	[0.0133]††
Augmented: adds female-student × covariate and × term interactions			
Female Student × Female Instructor	0.0361	0.0388	0.0444
	(0.0144)**	(0.0149)***	(0.0148)***
	[0.0144]††	[0.0148]†††	[0.0000]
F-test, female-student × covariate block = 0 (p-value, 1-way cluster)	0.000	0.017	0.006
F-test, female-student × covariate block = 0 (p-value, 2-way cluster)	0.000	0.000	n.a.
F-test, female-student × term block = 0 (p-value, 1-way cluster)	0.084	0.176	0.007
F-test, female-student × term block = 0 (p-value, 2-way cluster)	0.063	0.331	n.a.
N	3,018	3,018	3,017
Fixed Effects Included			
Term	Yes	No	No
Term×Course ×Days×Time	No	Yes	No
Class	No	No	Yes

Notes: The table tests whether the gender-match coefficient is identified by a functional-form artifact in which female students respond differently to other student-level covariates. The augmented sub-panel adds, on top of the baseline gender-match specification, the interaction of the female-student indicator with each of the eleven baseline student covariates listed in Section 3 (continuous covariates enter linearly, and categorical covariates enter as a full set of indicator interactions), and the interaction of the female-student indicator with the term fixed effect. The comparison sub-panel reports the gender-match coefficient from the baseline specification on the same sample for direct comparison. All specifications include the full set of student, instructor, and class controls. One-way clustered standard errors (by assignment cell) are shown in parentheses, and two-way clustered standard errors (by assignment cell and instructor) are shown in brackets. The joint F-tests at the bottom of each panel test the null that the entire female-student-covariate (or female-student-term) interaction block equals zero. Class fixed effects (column 3) absorb all section-level identifiers, and the female-student-covariate and female-student-term interactions remain identified within sections because both blocks vary at the student level. The label "n.a." indicates that the two-way cluster-robust variance matrix for the joint Wald test was not invertible under class fixed effects. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

Table A8: Leave-One-Out Jackknife: Gender-Match Coefficient Excluding Each Female Instructor

	Advanced Economics Course			Economics Major		
Excluded Instructor	$\hat{\delta}$	SE	[SE_{2w}]	$\hat{\delta}$	SE	[SE_{2w}]
Baseline (none excluded)	0.0619	(0.0289)		0.0241	(0.0131)	
Instructor A	0.0878	(0.0324)	[0.0282]	0.0279	(0.0150)	[0.0153]
Instructor B	0.0542	(0.0304)	[0.0319]	0.0192	(0.0135)	[0.0129]
Instructor C	0.0691	(0.0293)	[0.0318]	0.0262	(0.0134)	[0.0132]
Instructor D	0.0601	(0.0303)	[0.0344]	0.0252	(0.0139)	[0.0138]
Instructor E	0.0617	(0.0290)	[0.0311]	0.0241	(0.0132)	[0.0129]
Instructor F	0.0616	(0.0289)	[0.0308]	0.0240	(0.0132)	[0.0128]
Instructor G	0.0619	(0.0289)	[0.0310]	0.0243	(0.0132)	[0.0129]
Instructor H	0.0458	(0.0292)	[0.0273]	0.0159	(0.0123)	[0.0107]
Instructor I	0.0602	(0.0296)	[0.0317]	0.0256	(0.0135)	[0.0133]
Instructor J	0.0690	(0.0292)	[0.0325]	0.0275	(0.0134)	[0.0136]
Instructor K	0.0587	(0.0282)	[0.0349]	0.0252	(0.0141)	[0.0141]
Instructor L	0.0579	(0.0290)	[0.0307]	0.0217	(0.0130)	[0.0127]
Instructor M	0.0582	(0.0296)	[0.0316]	0.0258	(0.0133)	[0.0131]
Instructor N	0.0666	(0.0299)	[0.0322]	0.0262	(0.0134)	[0.0132]

Notes: Each row drops all student-section observations taught by the indicated female instructor and re-estimates the preferred assignment-cell specification. Instructors are labeled A through N to preserve anonymity. The baseline row reproduces the full-sample estimate from Table 3. The largest jackknife deviation from the baseline is 0.80 standard errors for advanced course-taking and 0.67 standard errors for major completion. Standard errors clustered by assignment cell in parentheses, and two-way clustered by assignment cell and instructor in brackets.

Table A9: Stability of Gender-Match Effects Across Instructor Partitions

	Tenure Status		Teaching Experience	
	Tenured/TT	IAS	Above Median	Below Median
	(1)	(2)	(3)	(4)
Panel A: Advanced Economics Course				
Difference in Gender Gaps	0.059	0.037	0.050	0.075*
	(0.048)	(0.045)	(0.042)	(0.041)
	[0.033]	[0.042]	[0.049]	[0.027]
N	1,626	1,392	1,451	1,567
Equality p (tenure)	0.630			
Equality p (experience)			0.850	
Panel B: Economics Major				
Difference in Gender Gaps	0.022	0.012	0.009	0.033*
	(0.022)	(0.018)	(0.021)	(0.019)
	[0.016]	[0.019]	[0.019]	[0.019]
N	1,626	1,392	1,451	1,567
Equality p (tenure)	0.711			
Equality p (experience)			0.616	

Notes: Each column reports the gender-match interaction coefficient (Difference in Gender Gaps) from the baseline specification restricted to sections taught by instructors in the indicated partition. Columns (1)–(2) partition by tenure status (tenured or tenure-track vs. instructional academic staff). Columns (3)–(4) partition by years of teaching experience (above vs. below the sample median). All specifications include assignment-cell fixed effects and the full set of student, instructor, and class controls. Standard errors clustered by assignment cell in parentheses, and two-way clustered by assignment cell and instructor in brackets. Equality p-values are from the triple-interaction coefficient Female Student $\times$ Female Instructor $\times$ Partition Indicator in the pooled sample with assignment-cell FE and one-way clustering. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Wild Cluster Bootstrap Inference

Specification	Original t	Conventional p	Bootstrap p	Clusters
Table 3, Panel A: Advanced Economics				
FE = Term	2.3823	0.0172	0.0606	27
FE = Assignment Cell	2.1533	0.0313	0.0730	27
FE = Class	2.4047	0.0162	0.0432	27
Table 3, Panel B: Economics Major				
FE = Term	1.6153	0.1062	0.1035	27
FE = Assignment Cell	1.8050	0.0711	0.0836	27
FE = Class	1.9636	0.0496	0.0662	27

Notes: Wild cluster bootstrap p-values using Webb 6-point weights with 9,999 replications, clustering at the instructor level (prof_id). The null hypothesis ($\beta_{\text{interaction}} = 0$) is imposed by estimating the restricted model (omitting the interaction term) and resampling the restricted residuals. "Original t" and "Conventional p" are from reghdfe with one-way clustering by assignment cell. "Bootstrap p" reports $p = (1 + \sum \mathbf{1}[|t_b^*| \geq |\hat{t}|])/(B+1)$. All specifications include student, professor, and class controls. The reported test is on the gender-match interaction female $\times$ first_sem_fem from the Table 3 baseline specifications.

Table A11: Gender-Match Effects in Introductory Economics Courses, Heterogeneity Across Student, Instructor, and Class Characteristics, One-Way Clustered Standard Errors (Appendix)

	Probability of Taking An Advanced Economics Course				Probability of Completing An Economics Major			
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A1: First-Generation College Students								
Gender Gap with Male Instructor	−0.1622***	(0.0312)	−0.1635***	(0.0318)	−0.0497***	(0.0165)	−0.0529***	(0.0179)
Gender Gap with Female Instructor	−0.0529	(0.0416)	−0.0463	(0.0438)	−0.0019	(0.0211)	−0.0011	(0.0226)
Difference in the Gender Gaps	0.1093**	(0.0495)	0.1172**	(0.0509)	0.0478*	(0.0270)	0.0518*	(0.0294)
Panel A2: Non-First-Generation College Students								
Gender Gap with Male Instructor	−0.1404***	(0.0201)	−0.1494***	(0.0209)	−0.0308***	(0.0119)	−0.0330***	(0.0120)
Gender Gap with Female Instructor	−0.0963***	(0.0261)	−0.0988***	(0.0269)	−0.0182	(0.0128)	−0.0185	(0.0135)
Difference in the Gender Gaps	0.0441	(0.0317)	0.0506	(0.0322)	0.0126	(0.0173)	0.0145	(0.0182)
p-value, equality across subgroups	0.229		0.245		0.308		0.329	
Panel B1: Low Math Placement								
Gender Gap with Male Instructor	−0.1443***	(0.0241)	−0.1418***	(0.0246)	−0.0414***	(0.0133)	−0.0436***	(0.0133)
Gender Gap with Female Instructor	−0.1043***	(0.0281)	−0.0969***	(0.0274)	−0.0259*	(0.0139)	−0.0207	(0.0133)
Difference in the Gender Gaps	0.0400	(0.0344)	0.0448	(0.0345)	0.0155	(0.0188)	0.0229	(0.0178)
Panel B2: High Math Placement								
Gender Gap with Male Instructor	−0.1465***	(0.0244)	−0.1597***	(0.0258)	−0.0348***	(0.0115)	−0.0375***	(0.0125)
Gender Gap with Female Instructor	−0.0666**	(0.0337)	−0.0728**	(0.0351)	−0.0023	(0.0141)	−0.0067	(0.0152)
Difference in the Gender Gaps	0.0799*	(0.0414)	0.0869**	(0.0428)	0.0325*	(0.0182)	0.0309	(0.0197)
p-value, equality across subgroups	0.434		0.432		0.516		0.765	
Panel C1: Native Instructor								
Gender Gap with Male Instructor	−0.1374***	(0.0210)	−0.1394***	(0.0220)	−0.0402***	(0.0111)	−0.0393***	(0.0118)
Gender Gap with Female Instructor	−0.0548*	(0.0285)	−0.0529*	(0.0292)	−0.0078	(0.0143)	−0.0093	(0.0151)
Difference in the Gender Gaps	0.0826**	(0.0368)	0.0865**	(0.0374)	0.0324*	(0.0185)	0.0300	(0.0197)
Panel C2: Foreign Instructor								
Gender Gap with Male Instructor	−0.1673***	(0.0300)	−0.1846***	(0.0318)	−0.0311**	(0.0151)	−0.0419***	(0.0132)
Gender Gap with Female Instructor	−0.1247***	(0.0374)	−0.1251***	(0.0377)	−0.0212*	(0.0124)	−0.0186	(0.0120)
Difference in the Gender Gaps	0.0426	(0.0461)	0.0595	(0.0470)	0.0099	(0.0193)	0.0233	(0.0177)
p-value, equality across subgroups	0.505		0.663		0.408		0.807	
Panel D1: Share Female Less than 40%								
Gender Gap with Male Instructor	−0.1280***	(0.0262)	−0.1279***	(0.0263)	−0.0309***	(0.0111)	−0.0277***	(0.0104)
Gender Gap with Female Instructor	−0.0416	(0.0305)	−0.0455	(0.0323)	−0.0011	(0.0139)	−0.0029	(0.0140)
Difference in the Gender Gaps	0.0864**	(0.0396)	0.0824**	(0.0402)	0.0298*	(0.0167)	0.0248	(0.0167)
Panel D2: Share 40% or More								
Gender Gap with Male Instructor	−0.1621***	(0.0299)	−0.1747***	(0.0308)	−0.0448***	(0.0156)	−0.0509***	(0.0150)
Gender Gap with Female Instructor	−0.1183***	(0.0325)	−0.1170***	(0.0328)	−0.0225*	(0.0126)	−0.0231*	(0.0125)
Difference in the Gender Gaps	0.0438	(0.0439)	0.0576	(0.0446)	0.0223	(0.0191)	0.0278	(0.0191)
p-value, equality across subgroups	0.486		0.694		0.757		0.901	
Panel E1: Introductory Microeconomics								
Gender Gap with Male Instructor	−0.1469***	(0.0218)	−0.1566***	(0.0229)	−0.0421***	(0.0135)	−0.0464***	(0.0130)
Gender Gap with Female Instructor	−0.0713**	(0.0329)	−0.0696**	(0.0332)	−0.0201*	(0.0113)	−0.0187	(0.0114)
Difference in the Gender Gaps	0.0756*	(0.0407)	0.0870**	(0.0407)	0.0220	(0.0173)	0.0276	(0.0169)
Panel E2: Introductory Macroeconomics								
Gender Gap with Male Instructor	−0.1451***	(0.0261)	−0.1463***	(0.0264)	−0.0305***	(0.0112)	−0.0305**	(0.0119)
Gender Gap with Female Instructor	−0.1048***	(0.0301)	−0.1066***	(0.0311)	−0.0026	(0.0182)	−0.0047	(0.0184)
Difference in the Gender Gaps	0.0403	(0.0378)	0.0396	(0.0377)	0.0279	(0.0212)	0.0258	(0.0221)
p-value, equality across subgroups	0.531		0.402		0.831		0.949	
Fixed Effects Included:								
Term × Course × Days × Time	Yes		No		Yes		No	
Class	No		Yes		No		Yes	

Notes: The table reports percentage point differences in the probability that a student takes an advanced economics course (columns 1–4) or majors in economics (columns 5–8). Each specification includes 3,018 observations. Columns differ by the fixed effects listed at the bottom of the table. The estimate groupings (e.g., Panels A1 and A2, B1 and B2, etc.) show estimates from separate regressions (five in total) based on the regression estimates presented in Appendix Tables A3 and A4. One-way clustered standard errors (by assignment cell) are shown in parentheses. All specifications include the full set of student, instructor, and class controls. Instructor and class controls, other than instructor experience, are absorbed by assignment-cell fixed effects or fully accounted for by class fixed effects. The p-value reported at the bottom of each panel pair tests the null that the gender-match effect (the Difference in the Gender Gaps) is equal across the two subgroups of the corresponding heterogeneity dimension. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels.

Table A12: Gender-Match Effects in Introductory Economics Courses, Heterogeneity Across Student, Instructor, and Class Characteristics

	Probability of Taking An Advanced Economics Course				Probability of Completing An Economics Major			
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A1: First-Generation College Students								
Gender Gap with Male Instructor	−0.1622***	(0.0283)	−0.1635***	(0.0272)	−0.0497***	(0.0167)	−0.0529***	(0.0182)
Gender Gap with Female Instructor	−0.0529	(0.0527)	−0.0463	(0.0545)	−0.0019	(0.0197)	−0.0011	(0.0224)
Difference in the Gender Gaps	0.1093*	(0.0588)	0.1172**	(0.0597)	0.0478*	(0.0265)	0.0518*	(0.0303)
Panel A2: Non-First-Generation College Students								
Gender Gap with Male Instructor	−0.1404***	(0.0185)	−0.1494***	(0.0227)	−0.0308***	(0.0106)	−0.0330***	(0.0114)
Gender Gap with Female Instructor	−0.0963***	(0.0253)	−0.0988***	(0.0268)	−0.0182	(0.0143)	−0.0185	(0.0158)
Difference in the Gender Gaps	0.0441	(0.0290)	0.0506	(0.0311)	0.0126	(0.0171)	0.0145	(0.0187)
p-value, equality across subgroups	0.214		0.228		0.300		0.351	
Panel B1: Low Math Placement								
Gender Gap with Male Instructor	−0.1443***	(0.0312)	−0.1418***	(0.0322)	−0.0414***	(0.0091)	−0.0436***	(0.0092)
Gender Gap with Female Instructor	−0.1043***	(0.0247)	−0.0969***	(0.0237)	−0.0259	(0.0166)	−0.0207	(0.0162)
Difference in the Gender Gaps	0.0400	(0.0367)	0.0448	(0.0385)	0.0155	(0.0187)	0.0229	(0.0180)
Panel B2: High Math Placement								
Gender Gap with Male Instructor	−0.1465***	(0.0284)	−0.1597***	(0.0282)	−0.0348***	(0.0126)	−0.0375***	(0.0137)
Gender Gap with Female Instructor	−0.0666*	(0.0384)	−0.0728*	(0.0407)	−0.0023	(0.0109)	−0.0067	(0.0119)
Difference in the Gender Gaps	0.0799*	(0.0476)	0.0869*	(0.0478)	0.0325**	(0.0161)	0.0309*	(0.0172)
p-value, equality across subgroups	0.493		0.477		0.455		0.728	
Panel C1: Native Instructor								
Gender Gap with Male Instructor	−0.1374***	(0.0180)	−0.1394***	(0.0164)	−0.0402***	(0.0107)	−0.0393***	(0.0114)
Gender Gap with Female Instructor	−0.0548	(0.0334)	−0.0529	(0.0342)	−0.0078	(0.0167)	−0.0093	(0.0174)
Difference in the Gender Gaps	0.0826**	(0.0365)	0.0865**	(0.0366)	0.0324*	(0.0195)	0.0300	(0.0204)
Panel C2: Foreign Instructor								
Gender Gap with Male Instructor	−0.1673***	(0.0236)	−0.1846***	(0.0260)	−0.0311***	(0.0094)	−0.0419***	(0.0124)
Gender Gap with Female Instructor	−0.1247***	(0.0318)	−0.1251***	(0.0342)	−0.0212**	(0.0089)	−0.0186*	(0.0099)
Difference in the Gender Gaps	0.0426	(0.0357)	0.0595	(0.0385)	0.0099	(0.0115)	0.0233	(0.0148)
p-value, equality across subgroups	0.433		0.618		0.328		0.796	
Panel D1: Share Female Less than 40%								
Gender Gap with Male Instructor	−0.1280***	(0.0263)	−0.1279***	(0.0260)	−0.0309***	(0.0107)	−0.0277***	(0.0097)
Gender Gap with Female Instructor	−0.0416	(0.0396)	−0.0455	(0.0418)	−0.0011	(0.0181)	−0.0029	(0.0193)
Difference in the Gender Gaps	0.0864*	(0.0468)	0.0824*	(0.0472)	0.0298	(0.0198)	0.0248	(0.0209)
Panel D2: Share 40% or More								
Gender Gap with Male Instructor	−0.1621***	(0.0311)	−0.1747***	(0.0303)	−0.0448***	(0.0124)	−0.0509***	(0.0125)
Gender Gap with Female Instructor	−0.1183***	(0.0283)	−0.1170***	(0.0300)	−0.0225**	(0.0109)	−0.0231*	(0.0118)
Difference in the Gender Gaps	0.0438	(0.0408)	0.0576	(0.0399)	0.0223	(0.0143)	0.0278*	(0.0156)
p-value, equality across subgroups	0.514		0.697		0.725		0.904	
Panel E1: Introductory Microeconomics								
Gender Gap with Male Instructor	−0.1469***	(0.0213)	−0.1566***	(0.0233)	−0.0421***	(0.0114)	−0.0464***	(0.0123)
Gender Gap with Female Instructor	−0.0713***	(0.0267)	−0.0696***	(0.0267)	−0.0201*	(0.0117)	−0.0187	(0.0119)
Difference in the Gender Gaps	0.0756**	(0.0309)	0.0870***	(0.0321)	0.0220	(0.0157)	0.0276*	(0.0156)
Panel E2: Introductory Macroeconomics								
Gender Gap with Male Instructor	−0.1451***	(0.0236)	−0.1463***	(0.0187)	−0.0305***	(0.0108)	−0.0305***	(0.0111)
Gender Gap with Female Instructor	−0.1048**	(0.0444)	−0.1066**	(0.0469)	−0.0026	(0.0200)	−0.0047	(0.0192)
Difference in the Gender Gaps	0.0403	(0.0510)	0.0396	(0.0496)	0.0279	(0.0221)	0.0258	(0.0217)
p-value, equality across subgroups	0.519		0.377		0.820		0.943	
Fixed Effects Included:								
Term × Course × Days × Time	Yes		No		Yes		No	
Class	No		Yes		No		Yes	

Notes: The table reports percentage point differences in the probability that a student takes an advanced economics course (columns 1–4) or majors in economics (columns 5–8). Each specification includes 3,018 observations. Columns differ by the fixed effects listed at the bottom of the table. The estimate groupings (e.g., Panels A1 and A2, B1 and B2, etc.) show estimates from separate regressions (five in total) based on the regression estimates presented in Appendix Tables A3 and A4. Two-way clustered standard errors (by assignment cell and instructor) are shown in parentheses. All specifications include the full set of student, instructor, and class controls. Instructor and class controls, other than instructor experience, are absorbed by assignment-cell fixed effects or fully accounted for by class fixed effects. The p-value reported at the bottom of each panel pair tests the null that the gender-match effect (the Difference in the Gender Gaps) is equal across the two subgroups of the corresponding heterogeneity dimension. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels.

Table A13: Heterogeneity in Gender-Match Effects for First- and Non-First-Generation Students (Dependent Variable: Probability to Taking Advanced Economics Course)

	Math Placement		Instructor Nativity		Share of Females Enrolled in Class		Course Content	
	Low	High	Foreign	US-Born	$\leq 40\%$	$> 40\%$	Micro	Macro
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Non-First-Generation Students								
Gender Gap Among Non-First-Generation Students Taught by Male Instructors	−0.1246	−0.1515	−0.1421	−0.1431	−0.1268	−0.1452	−0.1497	−0.1205
	(0.0298)***	(0.0310)***	(0.0388)***	(0.0254)***	(0.0267)***	(0.0371)***	(0.0249)***	(0.0332)***
	[0.0304]†††	[0.0282]†††	[0.0303]†††	[0.0215]†††	[0.0297]†††	[0.0366]†††	[0.0275]†††	[0.0328]†††
Gender Gap Among Non-First-Generation Students Taught by Female Instructors	−0.1060	−0.0649	−0.1060	−0.0970	−0.0562	−0.1132	−0.0930	−0.0966
	(0.0348)***	(0.0355)*	(0.0427)**	(0.0330)***	(0.0359)	(0.0354)***	(0.0345)***	(0.0362)***
	[0.0292]†††	[0.0309]††	[0.0300]†††	[0.0366]†††	[0.0435]	[0.0346]†††	[0.0357]†††	[0.0387]††
Difference in the Gender Gaps Among Non-First-Generation Students	0.0186	0.0866	0.0361	0.0461	0.0706	0.0320	0.0568	0.0239
	(0.0426)	(0.0441)**	(0.0531)	(0.0409)	(0.0421)*	(0.0492)	(0.0411)	(0.0461)
	[0.0387]	[0.0376]††	[0.0309]	[0.0407]	[0.0467]	[0.0479]	[0.0391]	[0.0534]
Panel B: First-Generation Students								
Gender Gap Among First-Generation Students Taught by Male Instructors	−0.1741	−0.1716	−0.2322	−0.1377	−0.1481	−0.1676	−0.1342	−0.1710
	(0.0453)***	(0.0533)***	(0.0601)***	(0.0364)***	(0.0519)***	(0.0400)***	(0.0408)***	(0.0547)***
	[0.0514]†††	[0.0608]†††	[0.0416]†††	[0.0264]†††	[0.0386]†††	[0.0350]†††	[0.0433]†††	[0.0419]†††
Gender Gap Among First-Generation Students Taught by Female Instructors	−0.0646	−0.0311	−0.1532	0.0400	0.0325	−0.1206	−0.0292	−0.0884
	(0.0454)	(0.0667)	(0.0485)***	(0.0613)	(0.0545)	(0.0543)**	(0.0600)	(0.0513)*
	[0.0525]	[0.0859]	[0.0341]†††	[0.0604]	[0.0537]	[0.0560]††	[0.0600]	[0.0618]
Difference in the Gender Gaps Among First-Generation Students	0.1095	0.1406	0.0789	0.1777	0.1807	0.0470	0.1050	0.0826
	(0.0624)*	(0.0818)*	(0.0752)	(0.0647)***	(0.0726)**	(0.0653)	(0.0679)	(0.0751)
	[0.0735]	[0.1032]	[0.0524]	[0.0599]†††	[0.0622]†††	[0.0649]	[0.0686]	[0.0798]
N	1,350	1,668	1,085	1,933	1,526	1,492	1,823	1,195

Notes: The table reports percentage-point differences in the probability that a student takes at least one advanced economics course after the introductory class. Each column presents results from a separate regression allowing instructor–student gender-match effects to vary jointly by first-generation status and the contextual dimension indicated (math placement, instructor nativity, classroom gender composition, or course type). All specifications include the full set of student, instructor, and class controls; instructor and class controls other than instructor experience are absorbed by assignment-cell fixed effects. One-way clustered standard errors (by assignment cell) are shown in parentheses; two-way clustered standard errors (by assignment cell and instructor) are shown in brackets. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

Table A14: Heterogeneity in Gender-Match Effects for First- and Non-First-Generation Students (Dependent Variable: Probability of Completing an Economics Major)

	Math Placement		Instructor Nativity		Share of Females Enrolled in Class		Course Content	
	Low	High	Foreign	US-Born	$\leq 40\%$	$> 40\%$	Micro	Macro
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Non-First-Generation Students								
Gender Gap Among Non-First-Generation Students Taught by Male Instructors	-0.0325 $(0.0165)^{**}$ $[0.0178]^{\dagger}$	-0.0281 $(0.0160)^{*}$ $[0.0131]^{\dagger\dagger}$	-0.0158 (0.0153) $[0.0086]^{\dagger}$	-0.0410 $(0.0153)^{***}$ $[0.0147]^{\dagger\dagger\dagger}$	-0.0212 $(0.0124)^{*}$ $[0.0129]$	-0.0376 $(0.0209)^{*}$ $[0.0225]^{\dagger}$	-0.0335 $(0.0167)^{**}$ $[0.0150]^{\dagger\dagger}$	-0.0269 $(0.0162)^{*}$ $[0.0152]^{\dagger}$
Gender Gap Among Non-First-Generation Students Taught by Female Instructors	-0.0298 (0.0196) $[0.0264]$	0.0033 (0.0174) $[0.0133]$	-0.0205 (0.0130) $[0.0126]$	-0.0213 (0.0197) $[0.0237]$	-0.0086 (0.0170) $[0.0227]$	-0.0295 $(0.0167)^{*}$ $[0.0199]$	-0.0280 $(0.0154)^{*}$ $[0.0186]$	-0.0079 (0.0226) $[0.0180]$
Difference in the Gender Gaps Among Non-First-Generation Students	0.0027 (0.0251) $[0.0317]$	0.0314 (0.0243) $[0.0183]^{\dagger}$	-0.0047 (0.0201) $[0.0119]$	0.0197 (0.0258) $[0.0284]$	0.0126 (0.0194) $[0.0240]$	0.0081 (0.0260) $[0.0280]$	0.0055 (0.0234) $[0.0241]$	0.0190 (0.0268) $[0.0216]$
Panel B: First-Generation Students								
Gender Gap Among First-Generation Students Taught by Male Instructors	-0.0437 $(0.0219)^{**}$ $[0.0139]^{\dagger\dagger\dagger}$	-0.0571 $(0.0219)^{***}$ $[0.0240]^{\dagger\dagger}$	-0.0639 $(0.0285)^{**}$ $[0.0212]^{\dagger\dagger\dagger}$	-0.0421 $(0.0197)^{**}$ $[0.0192]^{\dagger\dagger}$	-0.0358 $(0.0208)^{*}$ $[0.0198]^{\dagger}$	-0.0459 $(0.0233)^{**}$ $[0.0197]^{\dagger\dagger}$	-0.0616 $(0.0245)^{**}$ $[0.0233]^{\dagger\dagger\dagger}$	-0.0269 (0.0220) $[0.0227]$
Gender Gap Among First-Generation Students Taught by Female Instructors	-0.0043 (0.0287) $[0.0186]$	0.0008 (0.0312) $[0.0303]$	-0.0241 (0.0338) $[0.0257]$	0.0178 (0.0258) $[0.0227]$	0.0166 (0.0299) $[0.0301]$	-0.0142 (0.0311) $[0.0259]$	-0.0057 (0.0263) $[0.0297]$	0.0103 (0.0366) $[0.0497]$
Difference in the Gender Gaps Among First-Generation Students	0.0394 (0.0356) $[0.0207]^{\dagger}$	0.0578 (0.0384) $[0.0401]$	0.0398 (0.0424) $[0.0277]$	0.0599 $(0.0338)^{*}$ $[0.0323]^{\dagger}$	0.0524 (0.0357) $[0.0358]$	0.0317 (0.0395) $[0.0343]$	0.0559 (0.0362) $[0.0383]$	0.0372 (0.0426) $[0.0549]$
N	1,350	1,668	1,085	1,933	1,526	1,492	1,823	1,195

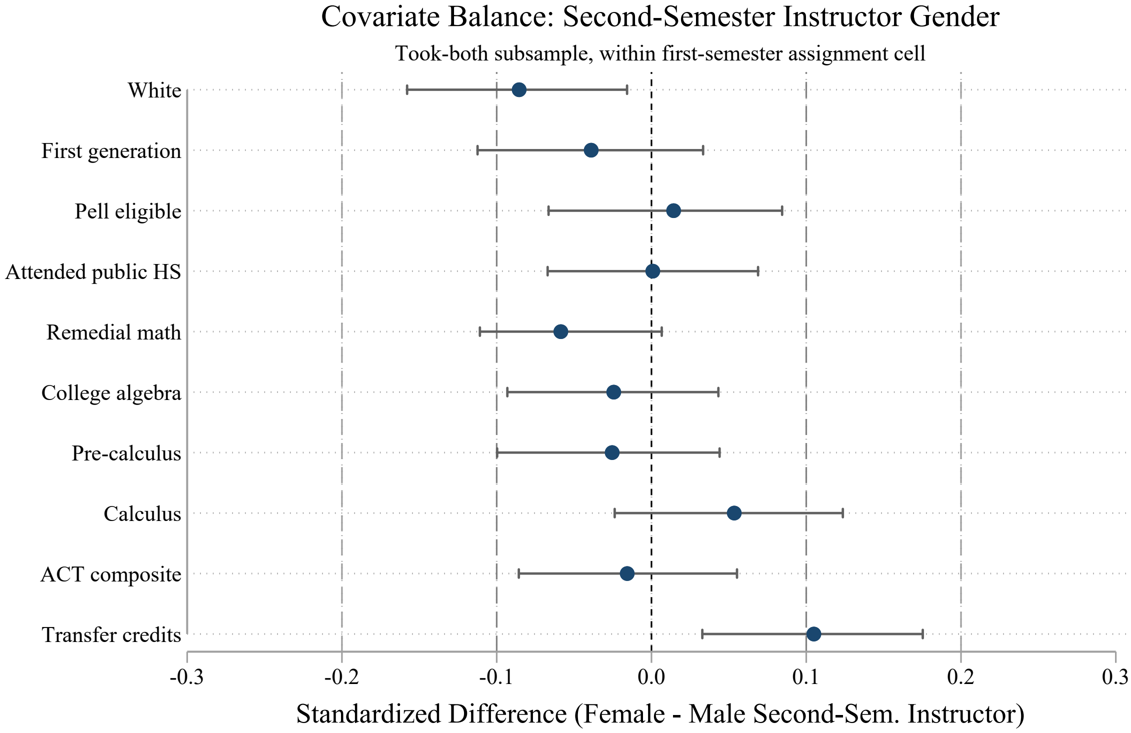
Notes: The table reports percentage-point differences in the probability that a student completes a major in economics. Each column presents results from a separate regression allowing instructor–student gender-match effects to vary jointly by first-generation status and the contextual dimension indicated (math placement, instructor nativity, classroom gender composition, or course type). All specifications include the full set of student, instructor, and class controls. Instructor and class controls other than instructor experience are absorbed by assignment-cell fixed effects. One-way clustered standard errors (by assignment cell) are shown in parentheses, and two-way clustered standard errors (by assignment cell and instructor) are shown in brackets. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

Table A15: Required Selection Magnitude to Mask a Hidden Cumulative-Exposure Effect

		Hypothesized true cumulative-exposure effect $= \alpha \cdot \widehat{\beta}_{\text{first}}$			
	Inputs	$\alpha = 0.25$	$\alpha = 0.50$	$\alpha = 0.75$	$\alpha = 1.00$
	(1)	(2)	(3)	(4)	(5)
Panel A: Probability of Taking Advanced Economics Courses					
$\widehat{\beta}_{\text{first}}$ (first-instructor effect)	0.0763				
$\widehat{\beta}_{\text{second}}$ (second-instructor effect)	-0.0094				
standard error	(0.0275)				
Selection-on-observables yardstick	-0.0054				
Hypothesized true effect		0.0191	0.0381	0.0572	0.0763
Required bias $\widehat{\text{bias}}_{\text{sel}}$		0.0285	0.0476	0.0667	0.0857
as a multiple of $\widehat{\beta}_{\text{second}}$ SE		1.04	1.73	2.43	3.12
Required δ		-5.24	-8.75	-12.25	-15.76
Panel B: Probability of Completing an Economics Major					
$\widehat{\beta}_{\text{first}}$ (first-instructor effect)	0.0339				
$\widehat{\beta}_{\text{second}}$ (second-instructor effect)	-0.0032				
standard error	(0.0135)				
Selection-on-observables yardstick	0.0006				
Hypothesized true effect		0.0085	0.0169	0.0254	0.0339
Required bias $\widehat{\text{bias}}_{\text{sel}}$		0.0116	0.0201	0.0286	0.0370
as a multiple of $\widehat{\beta}_{\text{second}}$ SE		0.86	1.49	2.12	2.74
Required δ		19.69	34.02	48.36	62.69

Notes: The exercise asks how strong selection on second-instructor gender would need to be to mask a hidden cumulative-exposure effect of size $\alpha \cdot \widehat{\beta}_{\text{first}}$, under selection-induced attenuation bias $\widehat{\beta}_{\text{second}} = \beta_{\text{true}} - \widehat{\text{bias}}_{\text{sel}}$. Column 1 reports the inputs to the calculation. $\widehat{\beta}_{\text{first}}$ is the headline first-instructor gender-match coefficient (Female Student $\times$ Female Instructor [1st]) estimated on the took-both subsample with the second-instructor block, the pre-treatment student-observables block, instructor and class controls, and assignment-cell fixed effects. $\widehat{\beta}_{\text{second}}$ is the female-student response to a female second-semester instructor, computed as the linear combination Female Instructor [2nd] + Female Student $\times$ Female Instructor [2nd] from the same regression. The selection-on-observables yardstick is the movement of $\widehat{\beta}_{\text{second}}$ when the pre-treatment student-observables block is added to the regression. Columns 2 to 5 sweep $\alpha \in \{0.25, 0.50, 0.75, 1.00\}$ and report the implied $\widehat{\text{bias}}_{\text{sel}} = \alpha \cdot \widehat{\beta}_{\text{first}} - \widehat{\beta}_{\text{second}}$, expressed as an absolute number, as a multiple of the standard error of $\widehat{\beta}_{\text{second}}$, and as a multiple δ of the yardstick (the Oster 2019 / Altonji, Elder, Taber 2005 proportionality concept). $\delta = 1$ means selection on unobservables is exactly as strong as selection on observables, $\delta > 1$ means stronger, and $\delta < 0$ means selection on unobservables would have to push in the opposite direction of selection on observables. The pre-treatment student-observables block contains ACT composite, math placement, transfer credits, first-generation status, race or ethnicity, Pell eligibility, and a public-high-school indicator. Standard errors clustered by assignment cell. When the yardstick is itself near zero (Panel B), the required δ explodes mechanically, indicating that selection on unobservables would have to be vastly larger than the measured selection on observables to mask the hypothesized hidden effect.

Figure A1: Standardized Imbalance — Second-Semester Instructor Gender, Took-Both Subsample



Notes: The figure plots standardized mean differences in pre-treatment student characteristics between students whose second-semester introductory-economics instructor is female and students whose second-semester instructor is male, on the took-both subsample ($N = 2{,}512$). Standardized differences are computed within first-semester assignment cells (term $\times$ course $\times$ day $\times$ time): each covariate is residualized on its first-semester cell mean, and the standardized difference is the residualized group-mean difference (Female 2nd $-$ Male 2nd) divided by the pooled standard deviation. Horizontal bars are 95% bootstrap percentile intervals based on 1,000 resamples (seed 20260415). Dashed vertical lines at ± 0.10 and ± 0.20 SD mark conventional thresholds for negligible and moderate imbalance. Nine of the ten covariates fall within the ± 0.10 band. The figure mirrors Figure 1 on the second-semester margin.

Appendix B — k-Means Clustering of Institutional Peers

To evaluate the representativeness of the study university (University of Wisconsin–La Crosse, UWL), we used institutional data from the U.S. Department of Education's *College Scorecard* to identify a set of peer universities. The analysis draws on data covering 2009–2019, the same period as the study's administrative records. For each institution, we averaged annual values over this decade to obtain stable measures of typical characteristics and to smooth out short-term variation.

The analytic sample was restricted to public, four-year, predominantly undergraduate universities located within the fifty U.S. states. Specifically, we included institutions classified as publicly controlled, offering at least bachelor's degrees, with undergraduate programs as their primary degree-granting activity. We excluded very small and very large institutions, defined as those enrolling fewer than 2,000 or more than 20,000 undergraduates, to focus on mid-sized universities most comparable to UWL. Observations with missing information on any clustering variable were dropped to ensure comparability across institutions.

The clustering used variables capturing institutional size, selectivity, student composition, affordability, and instructional resources. These included the admission rate, undergraduate enrollment, share of undergraduates who are White, share who are female, average annual net price (after grants and scholarships), percentage of undergraduates receiving Pell Grants, instructional expenditure per full-time-equivalent student, four-year graduation rate, student-to-faculty ratio, and a composite SAT-equivalent score. The SAT-equivalent variable combined reported SAT verbal, math, and writing midpoints when available, and converted ACT composites to the SAT scale using a concordance table. All variables were standardized to have mean zero and unit variance prior to clustering.

We implemented k-means clustering in Stata using the `cluster kmeans` command, specifying 10,000 random starts and up to 5,000 iterations for each solution to ensure convergence. To determine the number of clusters (k), we applied the elbow method following Makles (2012), plotting the total within-cluster sum of squares (WSS) for values of k between 2 and 20. The improvement in model fit declined sharply through $k = 5$ and flattened thereafter (Figure B1), indicating $k = 5$ as the optimal solution. Supplementary diagnostics based on the percentage reduction in WSS and curvature in $\log(\text{WSS})$ supported this conclusion. To test robustness, we also estimated cluster solutions for $k = 4$ and $k = 6$, and institutional memberships were highly consistent across these alternatives.

Under the $k = 5$ solution, UWL clusters with approximately 100 peer institutions that are best characterized as mid-sized, predominantly undergraduate public universities with moderate selectivity (admission rates near 70%), four-year graduation rates around 50–55%, average SAT-equivalent scores near 1150–1200, and Pell-receipt rates between 25–30%. These institutions are concentrated primarily across the Midwest and neighboring regions. Their aggregate profile closely mirrors that of UWL and represents the common segment of the U.S. public higher-education landscape.

Peer-group lists for the $k = 4$, $k = 5$, and $k = 6$ solutions are provided in Tables B1, B2, and B3 below. The analysis confirms that the study university belongs to a broad, representative group of mid-sized public institutions where most U.S. undergraduates encounter introductory economics. This benchmarking exercise situates our setting squarely within the common segment of the U.S. higher-education landscape, underscoring the external relevance of the results presented in the main text.

Table B1: List of Universities in UWL's Peer Group, $k = 4$

Angelo State University	Louisiana State University-Shreveport	University of Akron Main Campus
Arizona State University Digital Immersion	Louisiana Tech University	University of Alaska Anchorage
Arizona State University-Downtown Phoenix	Mansfield University of Pennsylvania	University of Arkansas at Little Rock
Arkansas State University	Marshall University	University of Arkansas-Fort Smith
Arkansas Tech University	McNeese State University	University of Baltimore
Armstrong State University	Midwestern State University	University of California-Merced
Auburn University at Montgomery	Minnesota State University Moorhead	University of Central Arkansas
Austin Peay State University	Minnesota State University-Mankato	University of Central Missouri
Bemidji State University	Minot State University	University of Central Oklahoma
Black Hills State University	Missouri Southern State University	University of Colorado Colorado Springs
Bloomsburg University of Pennsylvania	Missouri State University-Springfield	University of Louisiana at Lafayette
Boise State University	Montana State University	University of Louisiana at Monroe
Bridgewater State University	Montana State University Billings	University of Memphis
California State Polytechnic University-Humboldt	Morehead State University	University of Michigan-Dearborn
California State University-Chico	Murray State University	University of Michigan-Flint
California University of Pennsylvania	New Mexico State University-Main Campus	University of Missouri-St Louis
Central Washington University	Nicholls State University	University of Montevallo
Clarion University of Pennsylvania	North Dakota State University-Main Campus	University of Nebraska at Kearney
Cleveland State University	Northern Kentucky University	University of Nebraska at Omaha
Coastal Carolina University	Northern Michigan University	University of New Orleans
College of Staten Island CUNY	Northwest Missouri State University	University of North Alabama
Colorado Mesa University	Northwestern State University of Louisiana	University of North Carolina Asheville
Colorado State University Pueblo	Oakland University	University of North Carolina at Greensboro
Columbus State University	Old Dominion University	University of North Florida
Concord University	Pennsylvania State University-Penn State Abington	University of North Georgia
Delta State University	Pennsylvania State University-World Campus	University of Northern Colorado
East Stroudsburg University of Pennsylvania	Pittsburg State University	University of South Alabama
East Tennessee State University	Portland State University	University of South Carolina-Upstate
Eastern Kentucky University	Purdue University Fort Wayne	University of South Dakota
Eastern Michigan University	Purdue University-Calumet Campus	University of South Florida-St Petersburg
Eastern Oregon University	Purdue University-North Central Campus	University of Southern Indiana
Eastern Washington University	Radford University	University of Southern Maine
Edinboro University of Pennsylvania	Rhode Island College	University of Southern Mississippi
Emporia State University	SUNY College of Agriculture and Technology at Cobleskill	University of Toledo
Fairmont State University	Saginaw Valley State University	University of Washington-Bothell Campus
Farmingdale State College	Saint Cloud State University	University of Washington-Tacoma Campus
Ferris State University	Salem State University	University of West Florida
Fitchburg State University	Sam Houston State University	University of West Georgia
Florida Gulf Coast University	Shawnee State University	University of Wisconsin-Eau Claire
Fort Hays State University	Shepherd University	University of Wisconsin-Green Bay
Fort Lewis College	Slippery Rock University of Pennsylvania	University of Wisconsin-Oshkosh
Framingham State University	Sonoma State University	University of Wisconsin-Parkside
Frostburg State University	South Dakota State University	University of Wisconsin-River Falls
Georgia College & State University	Southeast Missouri State University	University of Wisconsin-Stevens Point
Georgia Southern University	Southeastern Louisiana University	University of Wisconsin-Stout
Henderson State University	Southeastern Oklahoma State University	University of Wisconsin-Superior
Indiana State University	Southern Oregon University	University of Wisconsin-Whitewater
Indiana University of Pennsylvania-Main Campus	Southern Utah University	Utah State University
Indiana University-East	Southwest Minnesota State University	Valdosta State University
Indiana University-Kokomo	Southwestern Oklahoma State University	Washburn University
Indiana University-Northwest	Stephen F Austin State University	West Liberty University
Indiana University-South Bend	Tarleton State University	West Texas A & M University
Indiana University-Southeast	Tennessee Technological University	West Virginia State University
Jacksonville State University	Texas A & M University-Corpus Christi	Western Kentucky University
Kean University	Texas Woman's University	Western Oregon University
Kutztown University of Pennsylvania	The Evergreen State College	Westfield State University
Lake Superior State University	The University of Montana	Wichita State University
Lamar University	The University of Tennessee-Chattanooga	Winona State University
Lander University	The University of Tennessee-Martin	Winthrop University
Lewis-Clark State College	The University of Texas Permian Basin	Worcester State University
Lock Haven University	The University of Texas at Tyler	Wright State University-Main Campus
Longwood University	Troy University	Youngstown State University

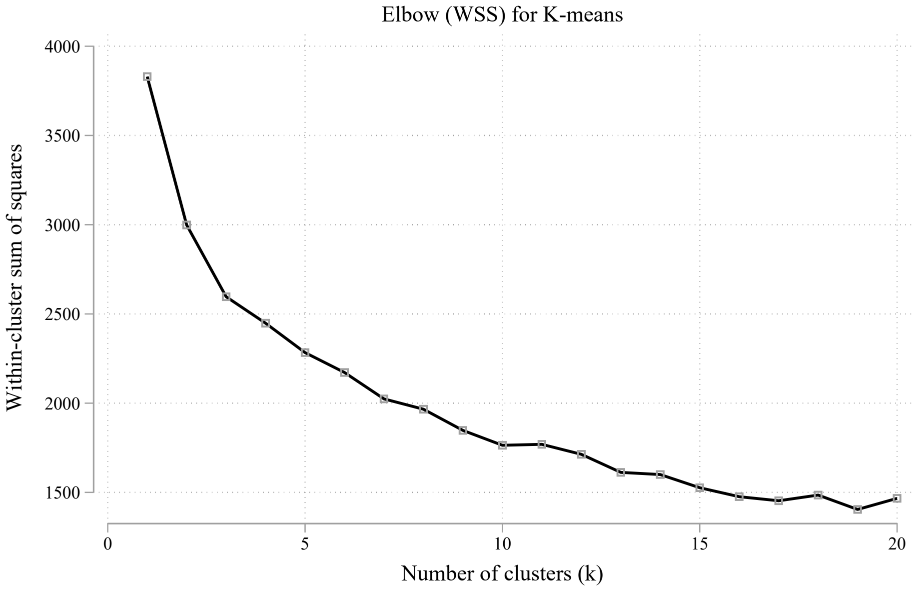
Table B2: List of Universities in UWL's Peer Group, $k = 5$

Appalachian State University	Montana State University	University of Maryland-Baltimore County
Ball State University	Montclair State University	University of Massachusetts-Dartmouth
Bloomsburg University of Pennsylvania	North Dakota State University-Main Campus	University of Massachusetts-Lowell
Boise State University	Oakland University	University of Minnesota-Duluth
Bowling Green State University-Main Campus	Oklahoma State University-Main Campus	University of Mississippi
Bridgewater State University	Old Dominion University	University of Nebraska at Omaha
California Polytechnic State University-San Luis Obispo	Pennsylvania State University-World Campus	University of Nebraska-Lincoln
California State University-Chico	Plymouth State University	University of Nevada-Reno
California University of Pennsylvania	Portland State University	University of New Hampshire-Main Campus
Central Connecticut State University	Radford University	University of North Carolina Wilmington
Central Michigan University	Ramapo College of New Jersey	University of North Carolina at Greensboro
Central Washington University	SUNY Brockport	University of North Florida
Christopher Newport University	SUNY College at Geneseo	University of Northern Iowa
College of Charleston	SUNY Oneonta	University of Oregon
East Stroudsburg University of Pennsylvania	SUNY at Fredonia	University of Rhode Island
Eastern Connecticut State University	Saint Cloud State University	University of South Dakota
Eastern Michigan University	Salisbury University	University of Toledo
Eastern Washington University	Shippensburg University of Pennsylvania	University of Washington-Bothell Campus
Florida Gulf Coast University	Slippery Rock University of Pennsylvania	University of Wisconsin-Eau Claire
Framingham State University	Sonoma State University	University of Wisconsin-Oshkosh
Frostburg State University	South Dakota State University	University of Wisconsin-Stevens Point
Georgia College & State University	Southern Illinois University-Edwardsville	University of Wisconsin-Stout
Georgia Southern University	State University of New York at Cortland	University of Wisconsin-Whitewater
Illinois State University	State University of New York at Oswego	Utah State University
Indiana University of Pennsylvania-Main Campus	Stockton University	West Chester University of Pennsylvania
James Madison University	Tennessee Technological University	Western Carolina University
Keene State College	The University of Texas at Dallas	Western Connecticut State University
Kutztown University of Pennsylvania	Towson University	Western Michigan University
Lock Haven University	Truman State University	Western Washington University
Longwood University	University of Akron Main Campus	Westfield State University
Miami University-Oxford	University of California-Santa Cruz	Winona State University
Millersville University of Pennsylvania	University of Colorado Colorado Springs	Worcester State University
Minnesota State University-Mankato	University of Idaho	Wright State University-Main Campus
Mississippi State University	University of Maine	
Missouri State University-Springfield	University of Mary Washington	

Table B3: List of Universities in UWL's Peer Group, $k = 6$

Appalachian State University	Plymouth State University	University of Illinois Springfield
Ball State University	Radford University	University of Maine
Bloomsburg University of Pennsylvania	Ramapo College of New Jersey	University of Mary Washington
Bowling Green State University-Main Campus	Rutgers University-Camden	University of Massachusetts-Dartmouth
Central Connecticut State University	SUNY Brockport	University of Minnesota-Duluth
Christopher Newport University	SUNY College at Geneseo	University of Missouri-Kansas City
Clarion University of Pennsylvania	SUNY College at Plattsburgh	University of Montevallo
College of Charleston	SUNY College at Potsdam	University of Nebraska at Kearney
Eastern Connecticut State University	SUNY Oneonta	University of New Hampshire-Main Campus
Eastern Illinois University	SUNY at Fredonia	University of North Carolina Asheville
Fitchburg State University	SUNY at Purchase College	University of North Carolina Wilmington
Framingham State University	Salem State University	University of Northern Iowa
Frostburg State University	Salisbury University	University of Pittsburgh-Johnstown
Georgia College & State University	Shippensburg University of Pennsylvania	University of Rhode Island
Indiana University of Pennsylvania-Main Campus	Slippery Rock University of Pennsylvania	University of South Dakota
Keene State College	South Dakota State University	University of Southern Maine
Kutztown University of Pennsylvania	Southern Connecticut State University	West Chester University of Pennsylvania
Longwood University	Southern Illinois University-Edwardsville	Western Carolina University
Mansfield University of Pennsylvania	State University of New York at Cortland	Western Connecticut State University
Millersville University of Pennsylvania	State University of New York at New Paltz	Western Illinois University
Montana State University	State University of New York at Oswego	Western Washington University
Montclair State University	Stockton University	Westfield State University
Oregon Institute of Technology	The College of New Jersey	William & Mary
Pennsylvania State University-Penn State Altoona	Towson University	William Paterson University of New Jersey
Pennsylvania State University-Penn State Berks	Truman State University	Winona State University
Pennsylvania State University-Penn State Harrisburg	University of Alabama in Huntsville	Winthrop University
Pennsylvania State University-World Campus	University of Idaho	Worcester State University

Figure B1: Elbow Plot: Within-Cluster Sum of Squares by Number of Clusters



Notes: The figure plots the total within-cluster sum of squares (WSS) from k-means clustering of U.S. public four-year predominantly undergraduate institutions for $k = 2$ to $k = 20$. The improvement in model fit declines sharply through $k = 5$ and flattens thereafter, indicating $k = 5$ as the optimal solution. Data are from the U.S. Department of Education College Scorecard, averaged over 2009–2019.

Appendix C — Permutation Test of Conditional Random Assignment

C.1 Setup and motivation

Our identification strategy treats students as conditionally randomly assigned to introductory sections within each assignment cell, defined by the interaction of term, course (introductory microeconomics or macroeconomics), day pattern (Monday/Wednesday/Friday, Tuesday/Thursday, or remote), and time block. Under that null, the female-instructor indicator is independent of any pre-assignment student characteristic conditional on the assignment cell. The standardized-difference balance check in Figure 1 is consistent with this null because no covariate exceeds conventional imbalance thresholds, but a balance figure alone does not bound the probability that the observed imbalance arose by chance under the null. Because the design assigns students within finely defined cells rather than across the full sample, the appropriate reference distribution is the within-cell distribution implied by random shuffling, not a parametric distribution under unconditional independence. The permutation test reported in this appendix delivers that reference distribution. Permutation inference for assignment-test settings of this form is recommended by Lehmann and Romano (2005) and Good (2006). Our design mirrors the within-stratum reshuffling tests applied to student-section assignment data by Carrell and West (2010) at the U.S. Air Force Academy and to teacher-assignment data by Lim and Meer (2017, 2020).

C.2 Permutation procedure

We restrict attention to the analytic sample of 3,018 student-section observations used to estimate the baseline gender-match specifications. For each baseline student covariate X_v we construct the within-cell residual $\tilde{X}_{v,i,c} = X_{v,i,c} - \bar{X}_{v,\cdot,c}$, where $\bar{X}_{v,\cdot,c}$ is the mean of X_v in the assignment cell c to which student i belongs. The observed within-cell standardized mean difference (SMD) is

$$\mathrm{SMD}_v^{\mathrm{obs}} = \frac{\bar{\tilde{X}}_v^F - \bar{\tilde{X}}_v^M}{\sqrt{\left(s_v^{F\,2} + s_v^{M\,2}\right)/2}},$$

where $\bar{\tilde{X}}_v^F$ and $\bar{\tilde{X}}_v^M$ are the female- and male-instructor sample means of $\tilde{X}_v$, and s_v^F and s_v^M are the corresponding sample standard deviations. We construct the permutation null distribution for $\mathrm{SMD}_v^{\mathrm{obs}}$ by drawing $B = 1{,}000$ random reassignments of the female-instructor indicator within each assignment cell. Each draw holds the cell-specific count of female-taught students fixed at its observed value and randomizes only the mapping between individual students and the female-versus-male instructor labels. Within a cell, this is equivalent to shuffling the values of the female-instructor indicator across students. Aggregated across cells, it is the within-stratum permutation distribution implied by exact conditional random assignment. For each draw $b = 1, \ldots, B$ we recompute $\mathrm{SMD}_v^{(b)}$ from the same residualization and pooling formulas, and we compute the empirical p-value

$$\hat{p}_v = \frac{1 + \sum_{b=1}^{B} \mathbb{1}\left[\,|\mathrm{SMD}_v^{(b)}| \geq |\mathrm{SMD}_v^{\mathrm{obs}}|\,\right]}{B + 1}.$$

The $(1 + \cdot)/(B + 1)$ adjustment is the conventional bound that prevents the empirical p-value from underflowing to zero. The random seed for the permutation draws is fixed in our replication code so that the table is exactly reproducible.

C.3 Results

Table C1 reports the observed within-cell SMD, the average SMD across the $1{,}000$ permutations, and the empirical p-value for each of the thirteen baseline student covariates that enter the regression specification. The null-mean column is uniformly close to zero. This confirms that the permutation procedure is correctly centered, since under random within-cell reassignment the standardized difference between female-taught and male-taught students should average to zero. The observed SMDs are likewise small in absolute magnitude, with all thirteen falling well within the ± 0.10 band that Figure 1 identifies as the threshold for negligible imbalance.

Twelve of the thirteen covariates fail to reject the conditional-random-assignment null at conventional levels, with empirical p-values above 0.10 and most well above 0.50. The single rejection is the indicator for White students, with an observed within-cell SMD of -0.050 and an empirical p-value of 0.038. The magnitude of this imbalance is small (under five percent of a pooled standard deviation and well below the 0.10 negligible-imbalance threshold), and it is the only rejection across thirteen simultaneous tests. Under the global null of conditional random assignment, the expected number of rejections at the five-percent level across thirteen independent tests is 0.65, so a single rejection is consistent with chance.

As a formal multiple-testing correction, the rightmost column of Table C1 reports Westfall-Young step-down familywise-error-rate-adjusted p-values from `wyoung`, computed from $1{,}000$ cluster-bootstrap replications of the same thirteen balance regressions clustered at the assignment cell (Westfall and Young, 1993). The Westfall-Young adjustment uses the joint resampling distribution of the test statistics across the thirteen covariates, so it accounts for correlation among baseline characteristics rather than treating the tests as independent. No covariate rejects the conditional-random-assignment null after this familywise correction. The single unadjusted rejection on the White indicator is therefore consistent with the design rather than evidence against it.

C.4 Implications for identification

Consistent with applied work on quasi-random scheduling settings (Lim and Meer, 2020), the body of the paper describes the assignment process as *plausibly quasi-random conditional on assignment cells*. The permutation test reported here supports that description rather than imposing it. Twelve of thirteen covariates fail to reject the conditional-random-assignment null, and the single rejection is consistent with chance, so the data do not contradict the design's identifying assumption that the female-instructor indicator is independent of pre-assignment student characteristics within an assignment cell.

Table C1: Permutation Test of Conditional Random Assignment

Baseline covariate	Observed SMD	Null mean SMD	Empirical p-value	WY-adj. p
Female student	0.0070	−0.0002	0.701	0.997
White	−0.0499	−0.0004	0.031**	0.969
First generation	0.0185	0.0005	0.358	0.997
Pell eligible	0.0093	−0.0000	0.675	0.997
Attended public HS	−0.0130	0.0007	0.512	0.974
Remedial math	−0.0176	−0.0003	0.357	0.974
College algebra	0.0152	−0.0003	0.516	0.990
Pre-calculus	0.0317	0.0004	0.162	0.866
Calculus	−0.0346	−0.0001	0.120	0.900
ACT composite	0.0012	−0.0004	0.965	0.997
Transfer credits	0.0065	0.0000	0.772	0.997
First-semester credits	−0.0175	−0.0017	0.439	0.985
Required advanced eco	0.0045	−0.0002	0.911	0.997

Notes: The table reports a permutation test of the conditional-random-assignment assumption that underpins the baseline gender-match estimates, with a Westfall-Young familywise-error-rate adjustment in the rightmost column. The analytic sample includes 3,018 student-section observations from the assignment-cell partition (term × course × day × time). For each baseline student covariate the column *Observed SMD* is the within-cell standardized mean difference between students taught by female and male instructors, computed as the residualized group-mean difference (Female – Male) divided by the pooled standard deviation. The column *Null mean SMD* is the average of the same statistic across 1000 permutations in which the female-instructor indicator is randomly reassigned across students within each assignment cell, holding the cell-specific count of female-taught students fixed. Under the conditional-random-assignment null this average should be approximately zero, and reading the column confirms that the permutation procedure is correctly centered. The *Empirical p-value* is the fraction of permutations whose absolute SMD is at least as large as the absolute observed SMD, computed using the conventional $(1+\sum_b \mathbb{1}[|\mathrm{SMD}_b| \geq |\mathrm{SMD}_{\mathrm{obs}}|])/(B+1)$ adjustment. The *WY-adj. p* column is the Westfall-Young step-down adjusted p-value from `wyoung` with 1000 cluster-bootstrap reps clustered at the assignment cell, applied to the family of 13 regressions of each baseline covariate on the female-instructor indicator with assignment-cell fixed effects, and it controls the familywise error rate across the simultaneous covariate tests (Westfall and Young, 1993). Significance markers refer to two-sided rejection at the indicated level. *, **, and *** denote $p < 0.10$, $p < 0.05$, and $p < 0.01$ respectively. The number of permutations is 1000 and the random seed is 20260417 (set in 01_setup.do).

Appendix D — Mediation Bounds and Predicted-Grade Heterogeneity for the Gender-Match Effect

This appendix reports the mediation-bounds exercise summarized in Section 5.3.2. The exercise replaces the sequential-ignorability assumption underlying the Imai-Keele-Yamamoto product-of-coefficients estimator with external literature anchors for the partial effect of grade on persistence. The IKY estimator reported in the main text is non-binding as an upper bound on the grade-mediated share in this sample because the sensitivity parameter at which the implied share collapses to zero is small ($\rho^* \approx 0.05$) relative to the canonical 0.20 to 0.40 range. Read in the other direction, that fragility is informative against a strongly grade-mediated reading. The literature-anchor bounds reported here use a different identifying assumption and reach a comparable upper bound.

D.1 Setup and decomposition

For each student, define the gender-match indicator $F = D_i \times D_{j(i)}$ using the student- and instructor-gender variables from equation 1. Let H be an indicator for earning a B or better in the first introductory economics course, and let Y be a downstream persistence outcome, either advanced-course enrollment or completion of an economics major. The reduced-form effect of F on Y is

$$\Delta Y = E[Y \mid F = 1] - E[Y \mid F = 0].$$

Following Imai et al. (2010), this reduced-form effect decomposes into a direct component and an indirect component operating through the grade mediator H:

$$\Delta Y \;=\; \underbrace{\Delta Y^{\text{direct}}}_{\text{non-grade component}} \;+\; \underbrace{\left(\frac{\partial Y}{\partial H}\right)\Delta H}_{\text{through grades}},$$

where $\Delta H = E[H \mid F = 1] - E[H \mid F = 0]$ is the treatment effect on grades and $\partial Y/\partial H$ is the marginal effect of earning a high grade on persistence holding F fixed. The reduced-form regression of Y on the gender-match interaction recovers ΔY, and a regression of H on the same interaction recovers ΔH. The partial $\partial Y/\partial H$ is not identified from our data. Plugging in plausible values from the literature traces out implied values of the controlled direct effect ΔY^{direct} and the share of the total reduced-form effect attributable to the indirect grade channel.

D.2 Why bounds rather than a clean estimate

Splitting ΔY into its two components requires knowledge of $\partial Y/\partial H$. Conditioning on H in the regression of Y on F does not recover $\partial Y/\partial H$ because H is post-treatment and is itself an outcome of F plus unobserved student factors. Recovering $\partial Y/\partial H$ from our data would require an instrument for grade that does not also predict the gender-match interaction. We lack such an instrument, as do existing studies of introductory-economics persistence at peer institutions. The bounds approach in this appendix sidesteps the identification problem. Instead of estimating $\partial Y/\partial H$ from our data, we use a grid of values from the literature on grade-to-major-choice effects in introductory economics, with a primary anchor near published estimates and grid points spanning the plausible range.

The same decomposition underlies the IKY bounds reported in the main text. The IKY bounds estimate $\partial Y/\partial H$ from our data under sequential ignorability and report the implied indirect share. That exercise is fragile in this sample because the sensitivity parameter at which the mediated share goes to zero is small ($\rho^* \approx 0.05$), making the IKY bound non-binding as a tight upper bound but informative against a strongly grade-mediated reading. Any economically trivial unobserved confounder of grade and persistence eliminates the mediated share. The literature-anchor approach reported here uses a different identifying assumption, namely that published estimates of grade-to-persistence effects in peer introductory-economics settings are informative for our population. The two exercises rest on different identifying assumptions and produce comparable upper bounds, with each piece of evidence reinforcing the qualitative conclusion that the grade channel does not dominate. The literature-anchor approach is itself subject to the caveat that the anchors are predominantly regression-discontinuity local effects rather than population-average partials of $\partial Y/\partial H$, an issue addressed in Section D.8.

D.3 Point estimates from the pooled-sample interaction specification

The primary specification is the pooled-sample assignment-cell regression from equation 1, repeated here for reference,

$$Y_i = \alpha + \beta D_i + \gamma D_{j(i)} + \delta\left(D_i \times D_{j(i)}\right) + X_i'\Theta + \eta_{c(i)} + \varepsilon_i,$$

where D_i and $D_{j(i)}$ are the student- and instructor-gender indicators, X_i is the standard control vector, $\eta_{c(i)}$ is the assignment-cell fixed effect, and δ is the gender-match parameter, our estimate of ΔY for $F = D_i \times D_{j(i)}$. The mediator regression replaces Y with the high-grade indicator H and uses the same right-hand side. Standard errors are clustered on assignment cell. We also report two-way clustered standard errors on (assignment cell, instructor) for the headline coefficients as a robustness check on inference.

The reduced-form interaction coefficient on advanced-course enrollment is 6.19 percentage points and is significant at the five-percent level under one-way and two-way clustering (Table D1). The coefficient on major completion is 2.41 percentage points and is significant at the ten-percent level under both schemes. The mediator interaction effect on the B-or-better indicator is 6.91 percentage points and is significant at the five-percent level (Table D2). Three features of the mediator estimates are worth noting. First, the interaction effect on the high-grade indicator is identical across the two outcomes by construction, since the mediator regression does not depend on Y. Second, the interaction-specification mediator estimate of 6.91 percentage points is smaller than the female-only level estimate of 11.56 percentage points (Table D3) because part of the female-instructor effect on grade is common across genders. The differential effect captured by the gender-match interaction is mechanically smaller than the level effect on female students alone. Third, the female-instructor effect on the grade is concentrated at the lower-pass-versus-fail margin (B-or-better and top tercile), with smaller and statistically insignificant effects at the AB-or-better and A-only thresholds. This pattern is consistent with an instructor effect that lifts students out of the lower tail rather than pushing students at the top.

D.4 Bounds grid

The decomposition above implies $\Delta Y^{\text{direct}} = \Delta Y - (\partial Y/\partial H) \cdot \Delta H$. For a grid of plausible anchor values of $\partial Y/\partial H$, plugging in the interaction-specification point estimates traces out implied values of the controlled direct effect (Tables D4 and D5).

The grid spans $\partial Y/\partial H$ from 0.025 to 0.30. At the smallest anchor the indirect channel through grade explains less than three percent of the total reduced-form interaction effect on advanced-course enrollment, and the implied controlled direct effect is essentially equal to the reduced-form coefficient. At the middle anchor of 0.10, the indirect channel explains about 11 percent of the total and the implied controlled direct effect is 5.5 percentage points. The remaining 89 percent is attributable to channels other than the measured grade. At the largest anchor of 0.30 the indirect channel caps at 33 percent of the total, and the implied controlled direct effect remains positive at 4.1 percentage points. No row in the grid drives the implied direct effect to zero or below, and no plausible literature anchor eliminates it. The qualitative reading is robust to anchor choice. The grade channel matters but does not dominate.

For the major-completion outcome the total interaction effect is smaller, at 2.41 percentage points, and the same mediator effect absorbs a larger share. At anchor 0.10 the indirect share is about 29 percent and the implied controlled direct effect is 1.7 percentage points. At anchor 0.30 the indirect channel can plausibly account for most of the small total effect on major declaration. Course-taking is mostly direct. Major declaration is more grade-mediated. The ordering is consistent with the major being a higher-stakes credentialing decision in which transcript quality carries more weight, while course-taking responds more to lower-cost factors such as encouragement and role-model exposure.

D.5 Literature anchors

The published literature on grade-to-major-choice effects in introductory economics provides the basis for the anchor grid. Owen (2010) uses a regression-discontinuity design to estimate the effect of receiving a higher introductory-economics grade on subsequent enrollment in economics, with attention to differential effects by gender. Main and Ost (2014) use a regression-discontinuity analysis to estimate the impact of letter grades on student effort, course selection, and major choice. McEwan et al. (2021) estimate grade sensitivity in the economics major at Wellesley College, a context that isolates female students. Antman et al. (2025) estimate the impact of relative grade signals on academic outcomes for students in introductory economics at the University of Colorado, the closest match to our institutional setting of an introductory-economics sequence at a U.S. public university. Most of these papers report estimates in the range $\partial Y/\partial H \in [0.05, 0.20]$, with some specifications reaching 0.30. We treat 0.10 as the modal anchor.

The grid in Table D4 maps each anchor to the implied controlled direct effect. Antman et al. (2025) is the closest setting to ours, so the corresponding row is the most defensible reading. The qualitative result is robust to anchor choice, since the implied controlled direct effect on advanced-course enrollment remains positive across the entire grid.

D.6 Robustness: female-only specification

The interaction specification quantifies how much of the differential gender-match effect for female versus male students runs through grade improvement. A complementary specification restricts the sample to female students and reads ΔY off the level coefficient on F. The two approaches answer different questions. The female-only approach gives a level effect for female students under weaker external-validity assumptions, since $\partial Y/\partial H$ only needs to be relevant for female students. The interaction approach matches the paper's headline specification and uses the full sample.

In the female-only specification ΔY is 5.88 percentage points on advanced-course enrollment and 1.99 percentage points on major completion (Table D3). The mediator effect on the B-or-better indicator is 11.56 percentage points and is significant at the five-percent level (Table D6). At anchor $\partial Y/\partial H = 0.10$, the implied controlled direct effect on advanced-course enrollment is 4.7 percentage points and the indirect share is about 20 percent. The female-only coefficients are larger than the interaction coefficients because they capture the level effect for female students rather than the differential over male students. The qualitative conclusion is the same as in the interaction specification, but the indirect share is larger because some of the female-instructor effect on grade is common to male students. The interaction differential strips that common component out.

D.7 Robustness: male-only specification

For completeness we report the male-only mirror exercise, in which $F = 1$ corresponds to an opposite-gender match.

The male sample shows a pattern qualitatively similar to the female sample but smaller in magnitude. ΔY on advanced-course enrollment is 3.63 percentage points, compared with 5.88 percentage points for female students (Table D7). ΔH on the B-or-better mediator is 7.83 percentage points, compared with 11.56 percentage points for female students (Table D8). None of the male reduced-form estimates is significant at conventional levels. Two readings are consistent with the data. Under the first, female instructors are modestly better instructors for both genders, with extra benefit for female students captured by the gender-match interaction. Under the second, the male positive coefficients are noise around zero and the female-only estimates capture the gender-match channel cleanly. The design cannot fully discriminate between these readings, but the qualitative result that the grade channel matters but does not dominate carries through under either.

D.8 Caveats

- **Additivity.** The decomposition assumes that the conditional expectation of Y is additive in F and H within the gender-by-instructor cells beyond the gender-match interaction we already include. An additional $F \times H$ interaction would prevent the decomposition from cleanly separating direct and indirect channels.
- **Single mediator.** The decomposition assumes that the introductory-course grade is the only mediator. If female-instructor exposure also operates through encouragement to attend office hours, take a second economics course, or invest in study habits, those alternative mediators are absorbed into the "direct" component. The implied ΔY^{direct} should therefore be read as everything not running through the measured grade rather than as a pure role-model effect. The discussion of identity, belonging, and instructor encouragement in Section 5.3.2 is consistent with this framing.
- **External validity of the anchor.** Importing $\partial Y/\partial H$ from external papers requires that the marginal effect of earning a high grade on persistence is similar in our population to the populations in those papers, and similar across genders for the interaction specification. Antman et al. (2025) is closest in setting. The other anchors give a sense of how sensitive the implied direct effect is to that assumption.
- **LATE versus ATE in the literature anchors.** Owen (2010), Main and Ost (2014), and McEwan et al. (2021) identify their estimates from regression-discontinuity designs at letter-grade thresholds on subsequent enrollment or major-completion outcomes. Owen (2010) uses the A/B cutoff. Main and Ost (2014) and McEwan et al. (2021) pool estimates across multiple cutoffs in the standard letter-grade scales used at their institutions (A/A–, A–/B+, B+/B, and so on). Those estimates are local average treatment effects of crossing a particular grade boundary, not the average partial effect of earning a B or better across the full distribution of introductory-course grades. The decomposition above treats $\partial Y/\partial H$ as the latter. Plugging an RD-LATE into the decomposition therefore treats a cutoff-specific marginal effect as the average partial across the support of H. The bias direction is ambiguous a priori. If compliers near a grade discontinuity have larger $\partial Y/\partial H$ than the population average, the literature-LATE plug-in overstates the indirect share and the bounds we report are conservative. If the population partial is larger than the threshold-LATE because the grade-to-persistence relationship is steeper away from the cutoff, the bounds understate the indirect share. Antman et al. (2025) identifies its effect from a relative-grade-signal information design rather than a sharp grade discontinuity, which partially mitigates this concern at the modal anchor.

- **Bounds, not point identification.** The exercise places quantitative bounds on a counterfactual without identifying it. The bounds indicate what would have to hold for the indirect channel to dominate the direct one, not whether that condition holds. The IKY sequential-ignorability bound reported in the main text is non-binding as a tight upper bound at the relevant sensitivity parameter ($\rho^* \approx 0.05$), but its fragility is itself informative against a strongly grade-mediated reading. The literature-anchor bounds reach a similar upper bound on the grade-mediated share under a different identifying assumption, subject to the LATE-versus-ATE caveat above.

D.9 Predicted-grade heterogeneity diagnostic

A grade-based mechanism predicts a preparation gradient in the persistence response. If female instructors operate by lifting course grades, the persistence response should covary with predicted preparation, since the grade-to-persistence margin is not flat across the support of H. An identity-based mechanism does not. We construct a predicted-grade index $\hat{G}(X)$ from pre-treatment covariates to test for such a gradient. We regress the introductory course grade on ACT composite score, math placement level, transfer credits, demographic indicators, and cohort fixed effects, estimated on the male-instructor subsample using five-fold cross-fitting to avoid overfitting bias. The resulting index explains about 24 percent of grade variation and correlates 0.44 with realized grades. Because $\hat{G}$ is a function of pre-treatment covariates, conditioning on it does not violate random assignment, and a formal balance check confirms that $\hat{G}$ is unrelated to instructor gender ($p = 0.45$). Interactions between $\hat{G}$ and the treatment indicator are therefore identified under the same assumptions as the main specification.

Table D9 reports the gender-match effect within quartiles of $\hat{G}$ among female students. The effect on advanced course-taking is positive in every quartile (Q1: 3.6, Q2: 8.4, Q3: 6.5, Q4: 4.7 percentage points), with the largest point estimate in Q2 and no monotone gradient in preparation. The effect on major completion follows a similar pattern, with point estimates between essentially zero and about 4 percentage points. The cross-quartile spread is small relative to the within-cell standard errors, and an F-test fails to reject equality of the four quartile effects for either outcome ($p = 0.80$ for advanced course-taking, $p = 0.55$ for major completion). The continuous triple-interaction specification yields the same conclusion, with the slope of the gender-match effect in $\hat{G}$ small and indistinguishable from zero. Among male students the quartile effects are flat and centered near zero, mirroring the baseline finding that male students do not respond to instructor gender. Panel C of Table D9 repeats the analysis with GPA as the outcome and finds the same flat quartile pattern.

The $\hat{G}(X)$ quartile pattern is flat for both outcomes. Read as a heterogeneity test, this rules out steep preparation gradients but not moderate heterogeneity. $\hat{G}(X)$ explains roughly 24 percent of grade variation, and the minimum detectable slope at 80 percent power, computed from the within-cell residual standard error of the continuous triple-interaction coefficient, translates to approximately 7.5 percentage points across the interquartile range of predicted grade for advanced course-taking and 4.1 percentage points for major completion. Because the average treatment effect among female students is roughly 5.9 and 2.0 percentage points for the two outcomes respectively (Appendix Table D3), the design can rule out gradients in which the effect more than doubles across the preparation distribution but cannot detect moderate heterogeneity. Alternative $\hat{G}$ constructions trained on male students rather than male-instructor students (Appendix Table D10, column 3) and using ACT composite alone (column 4) yield the same flat pattern.

D.10 Grade-to-persistence slope as a logical-consistency check

Two identified facts anchor the mechanism reading. Female instructors raise women's intro-course grades by 0.15–0.16 GPA points within assignment cell (Table 7), and the pre-treatment

$\hat{G}(X)$ heterogeneity in the gender-match effect is flat (Table D9). A uniform grade channel requires both a constant treatment-to-grade link and a constant grade-to-persistence link in $\hat{G}$. Panel C of Table D9 confirms the first link under random assignment. Appendix Table D11 tests the second by regressing persistence on GPA within $\hat{G}$ quartiles. The slope is near zero in the lowest $\hat{G}$ quartile and positive in the upper three, with the quintile pattern an inverted-U rather than a monotone gradient (Appendix Table D12). The pattern is robust to alternative $\hat{G}$ constructions (Appendix Table D13), a boundary-shift sample (Appendix Table D12, Panel C), and a leave-one-female-instructor-out jackknife (Appendix Table D14, Panel B). Identification of the second link as a causal grade-to-persistence elasticity requires sequential ignorability of grades, which the IKY sensitivity in Section 5.3.2 shows is fragile in this sample. We therefore do not interpret the non-flat slope as causal evidence against a uniform grade channel. The pattern remains a logical-consistency check rather than an identified test.

The non-flat grade-to-persistence slope nonetheless creates a logical tension with a uniform-grade reading. A uniform grade channel requires both (i) a constant treatment-to-grade link in $\hat{G}$, and (ii) a constant grade-to-persistence link in $\hat{G}$. Panel C of Table D9 supports (i). Appendix Table D11 is inconsistent with (ii) under sequential ignorability of grades. Reconciling the flat gender-match response across $\hat{G}$ with a non-flat grade-to-persistence slope under a grade channel therefore requires offsetting non-uniformities across (i) and (ii), a coincidence rather than a structural feature. An identity-and-belonging channel that does not require course success is consistent with the flat $\hat{G}$ pattern in the gender-match effect without requiring the offsetting coincidence, though it leaves the within-cell grade effect unexplained without additional structure linking belonging to course performance, for instance through differential engagement or effort responses. The two channels are not mutually exclusive and the design cannot isolate their separate contributions, but the patterns are easier to reconcile with a primarily identity-based reading than with a primarily grade-mediated one.

D.11 Summary

The gender-match effect on subsequent economics engagement is partly a grade-improvement story and mostly something else. The grade channel is real, with a precisely estimated 6.91-percentage-point first-stage on the gender-match interaction, and the grade lift partly explains downstream persistence. Under any plausible literature anchor for $\partial Y/\partial H$, however, there is also a meaningful non-grade direct effect, and for the course-taking outcome the non-grade channel is the dominant component. The predicted-grade heterogeneity diagnostic in Section D.9 reaches the same qualitative conclusion under a different identifying assumption, with the gender-match effect flat across pre-treatment quartiles of $\hat{G}(X)$ and inconsistent with steep preparation gradients of the kind a grade-based mechanism would predict. The mechanisms discussion in Section 5.3.2 uses these exercises to bound the contribution of the grade channel from above and to argue that the residual contains the channels emphasized in the main text (role modeling, encouragement, classroom climate, and identity and belonging) without overclaiming about which specifically fills it.

Table D1: Reduced-form regressions of Y on the gender-match interaction, pooled sample

	Advanced Economics Course	Economics Major
Female student × Female instructor	0.0619**	0.0241*
	(0.0289)	(0.0131)
N	3,018	3,018
R-squared	0.149	0.067

Notes: The table reports reduced-form regressions of two persistence outcomes on the gender-match interaction in the pooled estimation sample (N = 3,018). The two columns correspond to the two outcomes, an indicator for taking any advanced economics course (Adveco) and an indicator for completing an economics major (Eco major). The reported coefficient is on the gender-match interaction $F = D_i \times D_{j(i)}$, where D_i and $D_{j(i)}$ are the student- and instructor-gender indicators. Each regression also includes the main effects of D_i and $D_{j(i)}$, the standard student, instructor, and class control vector defined in Section 3, and assignment-cell fixed effects (term × course × day × time), all suppressed for brevity. One-way clustered standard errors on assignment cell are shown in parentheses. The two-way clustered standard error on (assignment cell, instructor) for the interaction coefficient is 0.0308 for advanced-course enrollment and 0.0128 for major completion. $*p < 0.10, **p < 0.05, ***p < 0.01$.

Table D2: Mediator regressions of H on the gender-match interaction, pooled sample

	H = grade $\geq$ B	H = grade $\geq$ AB	H = grade = A	H = top tercile
Female student $\times$ Female instructor	0.0691**	0.0321	0.0444*	0.0413
	(0.0334)	(0.0298)	(0.0228)	(0.0329)
N	3,018	3,018	3,018	3,018
R-squared	0.233	0.231	0.189	0.229

Notes: The table reports mediator regressions of four high-grade indicators on the gender-match interaction in the pooled estimation sample ($N = 3{,}018$). The reported coefficient is on the gender-match interaction $F = D_i \times D_{j(i)}$. The four columns differ only in the definition of the mediator H: an indicator for earning a B or higher in the introductory economics course (the headline definition used in the bounds exercise), an indicator for AB or higher, an indicator for an A, and an indicator for the top tercile of grade points within (term $\times$ course-number) cells. Each regression includes the main effects of D_i and $D_{j(i)}$, the standard student, instructor, and class controls defined in Section 3, and assignment-cell fixed effects, on the same right-hand side as Table D1. One-way clustered standard errors on assignment cell are shown in parentheses. The two-way clustered standard error on the headline B-or-better mediator is 0.0329. $*p < 0.10, **p < 0.05, ***p < 0.01$.

Table D3: Reduced-form regressions of Y on F, female students only

	Advanced Economics Course	Economics Major
Female instructor	0.0588	0.0199
	(0.0387)	(0.0161)
N	1,193	1,193
R-squared	0.263	0.127

Notes: The table reports reduced-form regressions of two persistence outcomes on the female-instructor indicator F, restricted to the female-student sub-sample ($N = 1{,}193$). The two columns correspond to the two outcomes, an indicator for taking any advanced economics course (Adveco) and an indicator for completing an economics major (Eco major). The reported coefficient is the level effect of female-instructor exposure on female students. Each regression includes the standard student, instructor, and class control vector defined in Section 3 and assignment-cell fixed effects (term $\times$ course $\times$ day $\times$ time), all suppressed for brevity. One-way clustered standard errors on assignment cell are shown in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$.

Table D4: Implied controlled direct interaction effect, Y = Advanced Economics Course, H = B-or-better, pooled sample

$\partial Y/\partial H$ anchor	ΔY	ΔH	implied ΔY^{direct}	share indirect
0.025	0.0619	0.0691	0.0602	0.028
0.050	0.0619	0.0691	0.0585	0.056
0.075	0.0619	0.0691	0.0567	0.084
0.100	0.0619	0.0691	0.0550	0.112
0.150	0.0619	0.0691	0.0516	0.167
0.200	0.0619	0.0691	0.0481	0.223
0.250	0.0619	0.0691	0.0447	0.279
0.300	0.0619	0.0691	0.0412	0.335

Notes: The table reports the implied controlled direct interaction effect of gender match on advanced economics course-taking for a grid of literature anchors on the partial effect of grade on persistence, $\partial Y/\partial H$. The mediator H is an indicator for earning a B or better in the introductory course. Each row corresponds to a different anchor value. Column 1 reports the assumed $\partial Y/\partial H$, the marginal effect of earning a high grade on advanced course-taking, holding the gender-match treatment fixed. Column 2 reports ΔY, the reduced-form interaction effect on advanced course-taking from Table D1. Column 3 reports ΔH, the mediator interaction effect on the high-grade indicator from Table D2. Column 4 reports the implied controlled direct effect, $\Delta Y^{\text{direct}} = \Delta Y - (\partial Y/\partial H) \cdot \Delta H$. Column 5 reports the share indirect, $(\partial Y/\partial H) \cdot \Delta H \,/\, \Delta Y$. The grid spans the plausible range of estimates from Owen (2010), Main and Ost (2014), McEwan et al. (2021), and Antman et al. (2025), with 0.10 as the modal anchor.

Table D5: Implied controlled direct interaction effect, Y = Economics Major, H = B-or-better, pooled sample

$\partial Y/\partial H$ anchor	ΔY	ΔH	implied ΔY^{direct}	share indirect
0.025	0.0241	0.0691	0.0224	0.072
0.050	0.0241	0.0691	0.0206	0.143
0.075	0.0241	0.0691	0.0189	0.215
0.100	0.0241	0.0691	0.0172	0.287
0.150	0.0241	0.0691	0.0137	0.430
0.200	0.0241	0.0691	0.0103	0.573
0.250	0.0241	0.0691	0.0068	0.717
0.300	0.0241	0.0691	0.0034	0.860

Notes: The table is the major-completion analog of Table D4 and reports the implied controlled direct interaction effect of gender match on completion of an economics major for the same grid of literature anchors on $\partial Y/\partial H$ and the same B-or-better mediator. Construction, sample, and column definitions are identical to Table D4. The reduced-form interaction effect ΔY is smaller for major completion (2.41 percentage points) than for advanced course-taking (6.19 percentage points), so the same mediator effect of 6.91 percentage points absorbs a larger share of the total. At the modal anchor of 0.10 the indirect share is about 29 percent and the implied controlled direct effect is 1.7 percentage points. At the upper-bound anchor of 0.30 the indirect channel absorbs most of the small total effect on major declaration.

Table D6: Mediator regressions of H on F, female students only

	H = grade $\geq$ B	H = grade $\geq$ AB	H = grade = A	H = top tercile
Female instructor	0.1156**	0.0447	0.0121	0.1024*
	(0.0536)	(0.0458)	(0.0366)	(0.0589)
N	1,193	1,193	1,193	1,193
R-squared	0.346	0.337	0.301	0.326

Notes: The table reports mediator regressions of four high-grade indicators on the female-instructor indicator F, restricted to the female-student sub-sample ($N = 1{,}193$). The four columns use four definitions of the high-grade indicator H: an indicator for earning a B or higher in the introductory course (the headline definition used in the bounds exercise), an indicator for AB or higher, an indicator for an A, and an indicator for the top tercile of grade points within (term $\times$ course-number) cells. Each regression includes the standard student, instructor, and class controls defined in Section 3 and assignment-cell fixed effects, all suppressed for brevity. One-way clustered standard errors on assignment cell are shown in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$.

Table D7: Reduced-form regressions of Y on F, male students only

	Advanced Economics Course	Economics Major
Female instructor	0.0363	0.0028
	(0.0352)	(0.0164)
N	1,820	1,820
R-squared	0.137	0.093

Notes: The table reports reduced-form regressions of two persistence outcomes on the female-instructor indicator F, restricted to the male-student sub-sample ($N = 1{,}820$). The two columns correspond to the two outcomes, an indicator for taking any advanced economics course (Adveco) and an indicator for completing an economics major (Eco major). The reported coefficient is the level effect of female-instructor exposure on male students, where $F = 1$ corresponds to an opposite-gender match. Each regression includes the standard student, instructor, and class control vector defined in Section 3 and assignment-cell fixed effects (term × course × day × time), all suppressed for brevity. One-way clustered standard errors on assignment cell are shown in parentheses. $*p < 0.10, **\,p < 0.05, ***p < 0.01$.

Table D8: Mediator regressions of H on F, male students only

	H = grade $\geq$ B	H = grade $\geq$ AB	H = grade = A	H = top tercile
Female instructor	0.0783	0.0357	0.0095	0.0843*
	(0.0512)	(0.0436)	(0.0272)	(0.0503)
N	1,820	1,820	1,820	1,820
R-squared	0.234	0.231	0.167	0.235

Notes: The table reports mediator regressions of four high-grade indicators on the female-instructor indicator F, restricted to the male-student sub-sample ($N = 1{,}820$). The four columns use four definitions of the high-grade indicator H: an indicator for earning a B or higher in the introductory course (the headline definition used in the bounds exercise), an indicator for AB or higher, an indicator for an A, and an indicator for the top tercile of grade points within (term $\times$ course-number) cells. Each regression includes the standard student, instructor, and class controls defined in Section 3 and assignment-cell fixed effects, all suppressed for brevity. One-way clustered standard errors on assignment cell are shown in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$.

Table D9: Heterogeneity in Gender-Match Effects by Predicted Grade

	Continuous Interaction		Quartile of Predicted Grade (Female Students Only)			
	Full Sample (1)	Female Only (2)	Q1 (Low) (3)	Q2 (4)	Q3 (5)	Q4 (High) (6)
Panel A: Advanced Economics Course						
Gender-match coefficient	0.0360	-0.0037	0.0361	0.0840	0.0651	0.0466
	(0.0743)	(0.0528)	(0.0464)	(0.0570)	(0.0530)	(0.0618)
	[0.0651]	[0.0651]	[0.0454]	[0.0448]†	[0.0705]	[0.0710]
p-value: $Q_1 = Q_2 = Q_3 = Q_4$			0.801			
p-value: $Q_4 = Q_1$			0.860			
N	3,018	1,196	1,196			
Panel B: Economics Major						
Gender-match coefficient	0.0382	0.0232	0.0085	0.0020	0.0415*	0.0205
	(0.0355)	(0.0290)	(0.0146)	(0.0301)	(0.0229)	(0.0264)
	[0.0380]	[0.0282]	[0.0192]	[0.0302]	[0.0225]†	[0.0195]
p-value: $Q_1 = Q_2 = Q_3 = Q_4$			0.549			
p-value: $Q_4 = Q_1$			0.664			
N	3,018	1,196	1,196			
Panel C: Intro-Course GPA (0–4)						
Gender-match coefficient	-0.0140	-0.1408	0.1451	0.2086*	-0.0023	0.1215
	(0.1078)	(0.1034)	(0.1167)	(0.1108)	(0.1164)	(0.1190)
	[0.0975]	[0.1244]	[0.1259]	[0.1241]	[0.1277]	[0.1405]
p-value: $Q_1 = Q_2 = Q_3 = Q_4$			0.339			
p-value: $Q_4 = Q_1$			0.839			
N	3,018	1,196	1,196			

Notes: Column (1) reports the triple-interaction coefficient Female Student × Female Instructor × $\hat{G}$ from the full sample. Column (2) reports the Female Instructor × $\hat{G}$ interaction among female students only. Columns (3)–(6) report the effect of a female instructor within each quartile of $\hat{G}$ among female students. Panel A reports effects on an indicator for taking any advanced (200+ level) economics course. Panel B reports effects on an indicator for majoring in economics. Panel C reports effects on the introductory-course grade in GPA points on the UW-La Crosse grading scale (F = 0, D = 1, C = 2, BC = 2.5, B = 3, AB = 3.5, A = 4). $\hat{G}(X)$ is the out-of-sample predicted intro-course GPA from a regression of grade on pre-treatment covariates, estimated on the male-instructor subsample using 5-fold cross-fitting. $\hat{G}$ is centered at its sample mean in columns (1) and (2). The mean predicted GPA rises from 2.18 in the lowest quartile to 2.51, 2.75, and 3.16 in the upper three quartiles. All specifications include assignment-cell fixed effects (term × course × day × time) and control for student, instructor, and class characteristics. One-way clustered standard errors (by assignment cell) are shown in parentheses; two-way clustered standard errors (by assignment cell and instructor) are shown in brackets. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering. *Sample comparison:* Female-only observations here (N = 1,196) differ slightly from the female-only reduced-form regression in Appendix Table D3 (N = 1,193). Both counts derive from the master estimation sample, and the three-observation gap reflects different singleton drops in two regression specifications with different right-hand sides.

Table D10: Robustness of Predicted-Grade Heterogeneity Analysis

	Main Specification (1)	Saturating Controls (2)	Male-Student Training (3)	ACT-Only Predictor (4)	Class FE (5)	All $X \times D$ Controls (6)
Panel A: Advanced Economics Course						
Q1 $\times$ Female Instructor	0.036	0.177	0.071	0.095	-0.023	0.160
	(0.046)	(0.263)	(0.048)	(0.053)	(0.062)	(0.290)
Q2 $\times$ Female Instructor	0.084	0.191	0.081	0.056	0.024	0.159
	(0.057)	(0.266)	(0.047)	(0.045)	(0.060)	(0.291)
Q3 $\times$ Female Instructor	0.065	0.142	0.063	0.050	-0.005	0.118
	(0.053)	(0.290)	(0.070)	(0.063)	(0.078)	(0.318)
Q4 $\times$ Female Instructor	0.047	0.079	0.036	0.015	0.000	0.069
	(0.062)	(0.307)	(0.060)	(0.055)	(0.000)	(0.335)
Panel B: Economics Major						
Q1 $\times$ Female Instructor	0.008	-0.090	0.003	0.006	-0.022	-0.115
	(0.015)	(0.144)	(0.015)	(0.023)	(0.028)	(0.154)
Q2 $\times$ Female Instructor	0.002	-0.103	0.027	0.029	-0.024	-0.150
	(0.030)	(0.151)	(0.025)	(0.022)	(0.036)	(0.162)
Q3 $\times$ Female Instructor	0.041	-0.074	0.011	0.027	-0.008	-0.122
	(0.023)	(0.150)	(0.028)	(0.024)	(0.034)	(0.161)
Q4 $\times$ Female Instructor	0.021	-0.124	0.040	0.020	0.000	-0.189
	(0.026)	(0.158)	(0.027)	(0.028)	(0.000)	(0.172)

Notes: Each cell reports the effect of a female instructor within a quartile of the predicted-grade index $\hat{G}$, estimated among female students. Column (1) reproduces the main specification (Table D9). Column (2) adds interactions between all student covariates and the female-instructor indicator. Column (3) trains $\hat{G}$ on male students rather than male-instructor students. Column (4) uses only ACT composite and cohort FEs to construct $\hat{G}$. Column (5) replaces assignment-cell FEs with class (section) FEs. Column (6) adds all $\hat{G}$-input covariates and their interactions with the treatment indicator as direct controls. All specifications include student, instructor, and class controls. Standard errors clustered by assignment cell in parentheses.

Table D11: Grade-Persistence Slope by Quartile of Predicted Grade (Female Students)

	Advanced Course (1)	Economics Major (2)
Q1 slope ($\hat{\beta}_1$)	-0.0181	0.0029
	(0.0224)	(0.0116)
	[0.0156]	[0.0094]
Q2 slope ($\hat{\beta}_1 + \hat{\beta}_{22}$)	0.1360***	0.0577***
	(0.0267)	(0.0212)
	[0.0256]†††	[0.0189]†††
Q3 slope ($\hat{\beta}_1 + \hat{\beta}_{23}$)	0.0907***	0.0520***
	(0.0270)	(0.0179)
	[0.0265]†††	[0.0167]†††
Q4 slope ($\hat{\beta}_1 + \hat{\beta}_{24}$)	0.0758**	0.0261*
	(0.0373)	(0.0134)
	[0.0271]†††	[0.0140]†
F-test (1-way): $\hat{\beta}_{22} = \hat{\beta}_{23} = \hat{\beta}_{24} = 0$	0.000	0.036
F-test (2-way): $\hat{\beta}_{22} = \hat{\beta}_{23} = \hat{\beta}_{24} = 0$	0.000	0.003
ρ^*, Q1 slope $\to 0$	-0.072	0.172
ρ^*, Q2 slope $\to 0$	0.684	0.460
ρ^*, Q3 slope $\to 0$	0.418	0.529
ρ^*, Q4 slope $\to 0$	0.320	0.323
ρ^*_{het} (flips any $\hat{\beta}_{2q}$)	0.365	0.255
N	1,196	1,196

Notes: Sample restricted to female students. Each column reports coefficient estimates from a single reghdfe regression of the stated persistence outcome on intro-course GPA, interactions of GPA with indicators for quartiles Q2–Q4 of the predicted-grade index $\hat{G}(X)$, main effects of the quartile indicators, the full student, instructor, and class control set, and assignment-cell fixed effects. The reported within-quartile slope is $\hat{\beta}_1 + \hat{\beta}_{2q}$, the partial derivative of the outcome with respect to GPA within quartile q. Identification of these slopes as causal persistence elasticities requires sequential ignorability of grade given controls, fixed effects, and treatment, the same assumption invoked for the grade-mediation bounds in Section 5.3.2. F-tests report the joint null that the three interaction coefficients are zero, i.e., that the slope is uniform across G-hat quartiles. ρ^* for each quartile reports the correlation between unobserved mediator and outcome determinants at which that quartile's slope is driven to zero, computed as $\rho^* = \hat{\beta}/(\sigma_M \sigma_Y)$ following Imai et al. (2010), where σ_M and σ_Y are residual standard deviations after partialling out the assignment-cell fixed effects and the full control set. ρ^*_{het} is the smallest $|\rho|$ across $q \in \{2, 3, 4\}$ at which any single interaction coefficient crosses zero, a conservative threshold for the value of omitted-variable confounding that would flip the heterogeneity test's conclusion. One-way clustered standard errors (by assignment cell) are shown in parentheses, and two-way clustered standard errors (by assignment cell and instructor) are shown in brackets. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

Table D12: Grade-Persistence Slope Heterogeneity: Alternative Bin Widths (Female Students)

	Tertiles of $\hat{G}$		Quintiles of $\hat{G}$		Boundary-shift check	
	Advanced	Major	Advanced	Major	Advanced	Major
1	0.0395	0.0255	-0.0276	0.0036	-0.0307	0.0051
	(0.0258)	(0.0157)	(0.0231)	(0.0132)	(0.0260)	(0.0150)
	[0.0234]	[0.0145]†	[0.0158]†	[0.0107]	[0.0188]	[0.0115]
2	0.1023***	0.0435***	0.1343***	0.0654***	0.1430***	0.0605*
	(0.0247)	(0.0155)	(0.0291)	(0.0232)	(0.0416)	(0.0306)
	[0.0227]†††	[0.0150]†††	[0.0281]†††	[0.0191]†††	[0.0354]†††	[0.0301]†
3	0.0787**	0.0395***	0.0826**	0.0289	0.0910***	0.0543***
	(0.0321)	(0.0143)	(0.0317)	(0.0186)	(0.0274)	(0.0179)
	[0.0296]††	[0.0189]††	[0.0368]††	[0.0174]	[0.0265]†††	[0.0170]†††
4			0.1277***	0.0600***	0.0726*	0.0275**
			(0.0316)	(0.0216)	(0.0387)	(0.0125)
			[0.0273]†††	[0.0221]††	[0.0280]††	[0.0128]††
5			0.0265	0.0143		
			(0.0458)	(0.0153)		
			[0.0351]	[0.0139]		
F-test (1-way)	0.265	0.701	0.000	0.081	0.001	0.112
F-test (2-way)	0.132	0.654	0.000	0.006	0.000	0.017
N	1,196	1,196	1,196	1,196	991	991

Notes. Sample restricted to female students. Each pair of columns reports the within-bin GPA-on-persistence slope under one of three alternative binnings of the predicted-grade index $\hat{G}$, separately for advanced-course enrollment and economics-major completion. Row labels 1–5 index bin position from lowest $\hat{G}$ (row 1) to highest. Tertiles (columns 2–3) use rows 1–3, quintiles (columns 4–5) use rows 1–5, and the boundary-shift check (columns 6–7) uses rows 1–4 on a sample that excludes women within ± 0.10 GPA units of the Q1/Q2 cutoff of $\hat{G}$. Each regression is of the stated outcome on intro-course GPA interacted with bin indicators, main effects of those indicators, the full student, instructor, and class control set, and assignment-cell fixed effects (same specification as Table D11). The F-test reports the p-value of the joint null that the within-bin slopes are equal across bins. One-way clustered standard errors (by assignment cell) in parentheses, two-way clustered standard errors (by assignment cell and instructor) in brackets. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

Table D13: Robustness of the Grade-Persistence Slope Heterogeneity (Female Students)

	Advanced Course (1)	Economics Major (2)
Panel A. ACT-only $\hat{G}$ (quartile heterogeneity)		
Q1 slope	0.0574	0.0292
	(0.0368)	(0.0199)
	[0.0289]†	[0.0193]
Q2 slope	0.0962***	0.0363***
	(0.0216)	(0.0135)
	[0.0215]†††	[0.0150]††
Q3 slope	0.0919***	0.0376**
	(0.0293)	(0.0175)
	[0.0348]††	[0.0208]†
Q4 slope	0.0503	0.0444***
	(0.0314)	(0.0157)
	[0.0295]†	[0.0184]††
F-test (1-way)	0.504	0.930
F-test (2-way)	0.049	0.889
Panel B. $\hat{G}$ trained on all male students		
Q1 slope	0.0305	0.0127
	(0.0286)	(0.0113)
	[0.0251]	[0.0117]
Q2 slope	0.1245***	0.0559***
	(0.0286)	(0.0201)
	[0.0288]†††	[0.0190]†††
Q3 slope	0.0458	0.0383**
	(0.0300)	(0.0185)
	[0.0315]	[0.0162]††
Q4 slope	0.1044***	0.0378***
	(0.0341)	(0.0128)
	[0.0291]†††	[0.0140]††
F-test (1-way)	0.097	0.127
F-test (2-way)	0.067	0.100
N (Panels A–B)	1,196	1,196

Notes. Sample restricted to female students. Panels A and B report within-quartile slopes from the same regression specification as Table D11, but using alternative predicted-grade indices. Panel A uses an $\hat{G}$ built from ACT composite plus cohort fixed effects alone. Panel B trains the $\hat{G}$ model on all male students rather than restricting to the male-instructor subsample. The F-test null in both panels is $\hat{\beta}_{22} = \hat{\beta}_{23} = \hat{\beta}_{24} = 0$. One-way clustered standard errors (by assignment cell) in parentheses, two-way clustered standard errors (by assignment cell and instructor) in brackets. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.

Table D14: Grade-Persistence Slope Heterogeneity: Q1-vs-Rest Tests and Leave-One-Instructor-Out Range (Female Students)

	Advanced Course (1)	Economics Major (2)
Panel A. Q1-vs-rest tests		
A.1. Drop Q1: $F(Q2 = Q3 = Q4)$ p (1-way)	0.232	0.314
A.1. Drop Q1: $F(Q2 = Q3 = Q4)$ p (2-way)	0.151	0.051
A.2. $I[Q1]\times$ GPA coef	-0.1176***	-0.0399***
	(0.0265)	(0.0144)
	[0.0198]†††	[0.0110]†††
Panel B. Leave-one-female-instructor-out range of p-values		
Quartile F p range (1-way)	[0.000, 0.000]	[0.001, 0.098]
Quartile F p range (2-way)	[0.000, 0.000]	[0.001, 0.008]
$I[Q1]\times$ GPA p range (1-way)	[0.000, 0.000]	[0.000, 0.030]
$I[Q1]\times$ GPA p range (2-way)	[0.000, 0.000]	[0.000, 0.005]

Notes. Sample restricted to female students. Panel A reports two versions of the Q1-vs-rest hypothesis. A.1 drops Q1 women and tests uniformity across Q2 to Q4. A.2 uses a binary indicator for Q1 interacted with GPA (baseline group is pooled Q2 to Q4). Panel B reports the range of p-values across 14 leave-one-female-instructor-out jackknife replications of the main quartile F-test from Table D11 and of the Panel A.2 binary $I[Q1]\times$ GPA coefficient. One-way clustered standard errors (by assignment cell) in parentheses, two-way clustered standard errors (by assignment cell and instructor) in brackets. The *, **, and *** denote statistical significance at the 10, 5, and 1% levels under one-way clustering, and †, ††, and ††† denote the same under two-way clustering.