Female Faculty Reduce Anxiety About STEM and Raise Academic Achievement Among Female Undergraduates

Robert Fairlie[1,2,3,4], Mridul Joshi[5], Saurabh Khanna[6], Prashant Loyalka[5,7], and Gagandeep Sachdeva[8]

[1]*Department of Economics, University of California, Los Angeles, Los Angeles, CA, USA*
[2]*National Bureau of Economic Research (NBER), Cambridge, MA, USA*
[3]*Institute of Labor Economics (IZA), Bonn, Germany*
[4]*CESifo Group Munich, Munich, Germany*
[5]*Graduate School of Education, Stanford University, Stanford, CA, USA*
[6]*Department of Communication Science, University of Amsterdam, Amsterdam, NH, NL*
[7]*Freeman Spogli Institute for International Studies, Stanford, CA, USA*
[8]*J-PAL South Asia, New Delhi, India*

Abstract

Despite rising college enrollment among women, gender disparities persist in STEM fields. We leverage a large-scale university setting in India, where undergraduates in computer science and electrical engineering programs are randomly assigned to instructors—a feature rarely feasible in higher education—to provide causal evidence on the effects of exposure to female faculty. Female students assigned to female faculty perform better on externally graded course exams and score higher on standardized tests administered two years later, indicating effects of female faculty that persist beyond the immediate classroom context. Female students also report lower anxiety about mathematics and science. The effects on anxiety and beliefs are consistent with exposure to identity-relevant cues that reduce psychological threat in STEM. Exposure to female faculty also shifts beliefs away from stereotypes about women's ability in mathematics and science, especially among male students. These findings suggest that repeated exposure to female faculty improves performance among female students and fosters more inclusive beliefs about women in STEM fields.

1 Introduction

Women earn more college degrees than ever before, yet remain markedly underrepresented in science, technology, engineering, and mathematics (STEM) fields such as computer science and engineering [1]. This gap limits economic opportunities and innovation while reinforcing social and economic inequality [2–4]. Stereotyping, performance expectations, and anxiety undermine women's participation and success in male-dominated STEM fields [5–8]. These factors are shaped by institutional environments, where faculty, mentors, and peers influence students' trajectories [9]. Understanding how these institutional factors interact with psychological barriers is key to addressing gender disparities in STEM [10].

Most causal studies exploring how women's experiences in STEM are shaped focus on peer mentorship ("horizontal exposure"), while the role of faculty ("vertical exposure") remains comparatively understudied, in part because large-scale, randomized faculty–student assignments are rare [11]. Without randomized interactions, it is difficult to identify causal effects without contamination from student selection. A substantial literature in social psychology demonstrates that female role models can reduce stereotype threat, bolster confidence, and support academic self-efficacy in STEM [12–16], and related work on role models and future-oriented aspirations suggests that exposure to female faculty can provide credible, attainable examples and broaden female students' beliefs about feasible STEM trajectories [17]. These studies offer important insight into mechanisms, though they often rely on smaller samples or controlled environments that may not fully capture a typical classroom setting. Complementary evidence from larger quasi-experimental studies in economics provides scale but yields mixed findings, with some showing neutral or even negative effects of female instructors on female students' grades and STEM persistence [18–22].

A complementary literature in primary and secondary education uses random or quasi-random teacher assignment to show that female teachers improve girls' test scores, engagement, mental health, social acclimation, and longer-run STEM outcomes [23–28]. Related literature shows that teachers' gender stereotypes widen gaps in girls' achievement, confidence, and educational choices [29–31]. These findings establish that both instructor identity and instructor beliefs shape student outcomes, though the postsecondary case is less settled, in part because random assignment of students to faculty is rare at that level.

To date, Carrell et al. [32] remains the only large-scale randomized study of faculty gender in STEM. Conducted at the U.S. Air Force Academy, it finds that female instructors improve female students' performance, with follow-up work linking such exposure to higher rates of STEM graduate study and careers [33]. Its randomized assignment process of Air Force cadets into classrooms provides rare causal evidence on the effects of exposure to female faculty. However, the study does not explore instructional practices or interpersonal dynamics that may underlie the observed effects.

Our study addresses this gap by bringing together a large-scale randomized design, a mass higher-education context, and rich psychological measures. We analyze data from 11 engineering colleges in India, a country that currently accounts for nearly a quarter of the global STEM graduate pipeline [34]. In India, women enter STEM programs, including computer science and engineering, at rates that exceed those in many higher-income countries. In 2021–22, women accounted for 29.1% of undergraduate enrollment in Engineering & Technology, 34.2% in Computer Engineering, and 37.6% in Information Technology & Computer Sciences [35]. These shares exceed the corresponding U.S. shares of bachelor's degrees awarded to women in engineering and computer and information sciences [36], yet enrollment need not translate mechanically into representation in technical jobs. Female labor-force participation in India remains far below male participation [37], and women account for only 18.6% of personnel directly engaged in R&D activities [38]. This gap between enrollment and technical employment makes the undergraduate period important for studying

inputs that shape later participation in technical fields. During these years, students accumulate human capital and academic records; they also update beliefs about their own ability, the costs and returns to effort, and their fit in technical fields. Studying female faculty and students in this setting allows us to examine whether representation in the classroom changes both academic performance and the beliefs students carry into later technical training and work.

The setting of our study provides randomized variation in exposure to female faculty in a high-stakes educational environment. This variation arises from an institutional feature that is uncommon in higher education research: students follow fixed curricula and are assigned to faculty-taught classrooms through a computerized procedure. This design allows us to cleanly identify the causal effect of exposure to female faculty in real classrooms, without the self-selection of students into courses or the awareness of participation in an experiment—factors that have posed challenges for identification in much of the prior literature.

We use this randomized variation in faculty gender to study both short-term academic outcomes and longer-run skill formation, as well as changes in students' psychological states and beliefs. Drawing on our large-scale data collection effort [39], we combine externally graded course marks and standardized mathematics and science assessments administered up to two years later with validated measures of STEM-related anxiety, confidence, and gender beliefs. This design allows us to study how identity-relevant cues in the learning environment may reduce anxiety about STEM, and accumulate into durable gains in human capital. In doing so, our study bridges two influential literatures—"teacher-like-me" designs in economics and social-psychological research on identity threat, belonging, and role models—while providing policy-relevant evidence on the effects of female faculty in a system that is central to the global STEM pipeline.

Here, we show that female students earn 0.083 standard deviations (2.64 course-rank percentile points) higher grades when taught by female faculty (95% CI: 0.021 to 0.146 SD; $P = 0.01$), with positive, statistically significant estimates among female students with lower prior achievement, lower STEM confidence, or lower commitment to their field of study at program entry. Over two years, a 10 percentage point increase in female faculty exposure raises female students' standardized test scores by 0.033 standard deviations relative to male students (95% CI: 0.004 to 0.063 SD; $P = 0.03$)—closing 18 percent of the gender gap in academic skills observed at program entry. Female faculty also reduce STEM-related anxiety by 0.053 standard deviations (95% CI: 0.022 to 0.084 SD; $P < 0.001$) and shift gender beliefs away from gender-based stereotypes, particularly among male students.

Conceptually, we interpret these patterns through a belief-based framework in which faculty gender acts as an identity-relevant signal. Exposure to a female instructor offers a credible and often novel cue about who belongs and who can succeed in STEM, reducing psychological threat, as reflected in lower STEM-related anxiety. These signals matter most for students who are more uncertain about their academic fit and thus more likely to revise their beliefs when presented with affirming cues. In line with this mechanism, we find large effects for students with lower commitment to their field of study. Gains in academic performance extend to standardized assessments

administered outside the original classroom context, suggesting that reductions in identity-related threat are associated with durable improvements in achievement.

2 Results

We first describe the research design, analytic sample, and random assignment. We then present the empirical results in five steps. First, we estimate the effect of female faculty on female students' course grades, leveraging random assignment of students to instructors. We then examine whether cumulative exposure to female faculty over the first two years improves performance on independently administered standardized mathematics and science assessments. Next, we examine course-grade effects among students with lower commitment to their field of study at program entry, motivated by the hypothesis that female faculty may provide a stronger signal of belonging for students who are less certain they fit in technical fields. We then examine effects on STEM-related anxiety and confidence to shed light on potential mechanisms underlying the academic outcomes. Finally, we present additional analyses that assess alternative explanations and broader classroom channels, including teaching practices, faculty–student interactions, faculty beliefs and credentials, and male peers' gender beliefs.

Research design and sample

This study uses randomized student assignment to faculty-taught sections within courses (hereafter "classrooms") across computer science and electrical engineering programs in 11 undergraduate engineering colleges in India, providing a causal estimate of the effects of female faculty on student outcomes. The sample includes 1,793 students (1,216 male, 577 female) who, over the first two years of their program, are randomly assigned across 1,084 classrooms taught by 431 instructors. Supplementary Fig. 1 summarizes the sample construction. Within each course, students are randomly assigned to classrooms, preventing selection based on faculty gender or peer composition. This design addresses the concern that female students who might benefit most from female faculty could also be more likely to choose their courses. Students follow a fixed curriculum in the first two years, ensuring that faculty or peer characteristics do not influence course selection.

Our empirical strategy leverages random assignment in two ways. First, we estimate the immediate effect of being assigned to a female instructor on course grades, using student and classroom fixed effects to absorb stable differences in student performance and all factors common to the same course section. This ensures that the effect of female faculty is identified from female–male grade differences within randomly assigned classrooms. Because the unit of analysis is the student–course pair, each student contributes multiple observations, increasing the effective sample size and improving statistical precision. Second, we examine longer-term impacts by assessing how variation in the proportion of courses taught by female faculty over two years influences academic and psychological outcomes. Because instructor assignment is random, differences in the amount of exposure to female faculty arise by chance rather than from student choices, so these estimates reflect the

causal effect of exposure. Further details on estimation procedures and model specifications are provided in Methods.

To estimate these effects, we constructed a dataset combining administrative records, standardized tests, and faculty and student surveys, linking faculty data to the students they taught. Administrative data track course enrollments, instructor assignments, and course grades, enabling comprehensive monitoring of students' academic progress. Because raw grades are not directly comparable across colleges or courses, we measure course grades using within-course ranks rather than raw grades, constructed across all classrooms taught by different faculty in the same course, college, and semester. A randomly selected 50 percent of students took standardized academic skills tests in mathematics and science (administered by the research team) at program entry and again two years later to measure learning gains. Of these 896 students, 867 completed the mathematics assessment and 885 completed the science assessment at the end of two years. Psychological outcomes, including STEM-related confidence and STEM-related anxiety, were constructed from multiple survey items and combined into indices using the inverse covariance matrix approach to improve reliability by accounting for correlations between related responses [40]. Gender beliefs were measured on a five-point scale and dichotomized as either stereotypical or non-stereotypical. A complete description of how outcomes were constructed is provided in Methods. High participation rates of approximately 95 percent in both survey rounds enhance data reliability [39].

Design validity and sample characteristics

Female and male students in our sample differ in key academic and demographic indicators (Supplementary Table 1). Women make up 32 percent of the sample, and on average, score about 0.20 SD (95% CI: 0.064 to 0.312 SD; $P = 0.003$) lower than men on the academic skills test administered by the research team at the start of their program. Female students, however, earn higher average course grades than male students in the realized college-grade data. Among students taught by male faculty, female students' average course grades are approximately 0.36 SD (11.42 course-rank points) higher than male students' average course grades. Female students are 8.5 percentage points (95% CI: 3.5 to 13.4 percentage points; $P < 0.001$) more likely to have a college-educated father and 5.7 percentage points more likely to have a college-educated mother (95% CI: 0.9 to 10.5 percentage points; $P = 0.02$). Female students are also 8.7 percentage points more likely to report that they would choose the same branch or department if choosing again (95% CI: 5.2 to 12.2 percentage points; $P < 0.001$). Female students report slightly lower STEM-related confidence and slightly higher STEM-related anxiety, but neither difference is statistically significant.

Tests of balance in student characteristics provide a check on whether students were randomly assigned across classrooms (Supplementary Table 2). We find no statistically significant differences across a range of student characteristics [41]. In addition, we perform two complementary statistical tests based on resampling. Within each unique course offering (a department-course combination, which in our data corresponds to a single institution and term), we reassign students across classrooms while preserving classroom sizes and compare the observed allocation to the

resulting randomization distribution. For continuous baseline covariates we report a Kolmogorov–Smirnov test of the uniformity of classroom-level empirical p-values. For binary baseline covariates we report an omnibus permutation test of between-classroom dispersion (Supplementary Table 3 and Supplementary Fig. 2). None of the reported permutation tests rejects its null hypothesis at the 5% level.

Female faculty comprise 34 percent of the sample (Supplementary Table 4). There are no statistically significant differences between female and male faculty in academic rank, PhD status, years of experience, holding a degree from an elite college, or eligibility for caste-based affirmative action.

Course grades and academic skills

Female students perform better in externally graded STEM courses when taught by female faculty, with no statistically significant effect on male students' performance. Course performance is measured using within-course ranks, constructed among students enrolled in the same course, college, and semester across randomly assigned classrooms. We standardize these ranks for the main analysis. On this measure, female students earn 0.083 standard deviations (2.638 course-rank percentile points) higher grades when taught by female faculty (95% CI: 0.021 to 0.146 SD; $P = 0.01$; Table 1).

Column 1 of Table 1 reports estimates from our primary within-classroom specification, which includes student fixed effects and classroom fixed effects. Student fixed effects absorb stable differences in students' average performance across courses, while classroom fixed effects absorb all factors common to students in the same course section, including the instructor, course, institution, and semester. The coefficient on female faculty $\times$ female student therefore estimates whether the female–male grade difference within a classroom is more favorable to female students when the instructor is female.

Because classroom fixed effects absorb the main effect of instructor gender, the specification in Table 1 cannot separately estimate the effect of female faculty on male students. We therefore report the male-student estimate in Supplementary Table 5, using an auxiliary specification that omits classroom fixed effects, includes course fixed effects, and adds faculty controls where indicated. The estimated effect on male students is not statistically significant (0.004 SD; 95% CI: -0.049 to 0.057; $P = 0.89$; Supplementary Table 5). As an interpretive aid, we combine observed mean course ranks in male-taught classes with the auxiliary estimates in Supplementary Table 5. In course-rank percentile points, male students taught by male faculty average 47.375 and female students taught by male faculty average 58.797. The corresponding implied values in female-taught classes are 47.502 for male students and 61.151 for female students.

The effect of female faculty on female students' grades is robust across a range of alternative specifications. We continue to find a positive effect when restricting the sample to non-elite colleges, addressing concerns that the effect might be driven by highly selective institutions. It also holds when limiting the analysis to science, mathematics, and engineering courses, ruling out the

influence of a small number of non-STEM courses present in the curriculum. The effect persists when weighting classrooms equally, addressing concerns about differential class sizes or potential distortions from errors in matching students to faculty. Finally, the result is consistent when using alternative scaling of the course-grade outcome, confirming that our findings are not sensitive to how student achievement is measured (Supplementary Table 6).

We next examine whether performance gains extend beyond the classroom, reflecting durable improvements in academic skill. Course grades may capture both effort and short-run responses to exposure to female faculty in the respective course [42]. To assess whether these benefits persist and extend beyond contexts where female faculty are present, we analyze performance on a standardized assessment of academic skills at the end of two years. Over two years, a 10 percentage point increase in female faculty exposure improves female students' standardized mathematics and science scores by 0.033 SD (95% CI: 0.004 to 0.063 SD; $P = 0.03$) relative to male students, closing 18 percent of the gender gap in academic skills (Table 1).

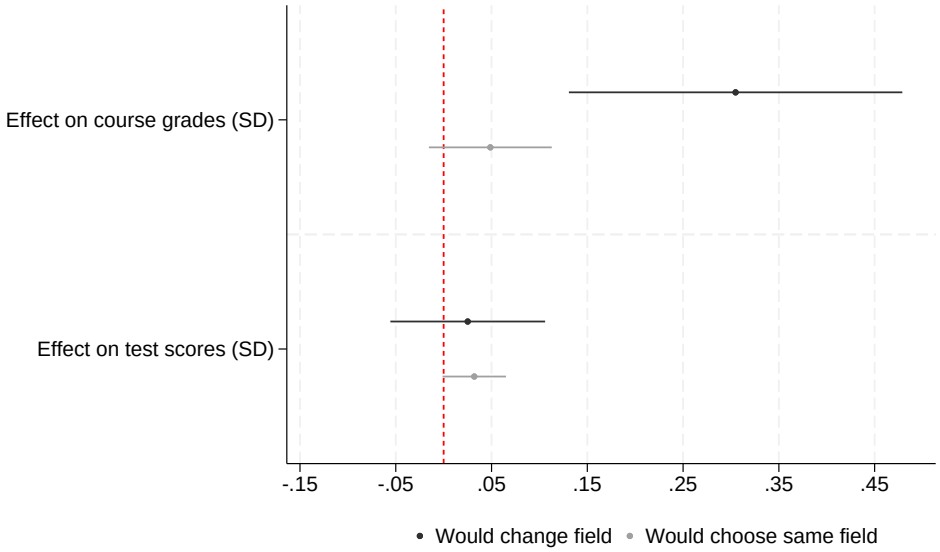
Figure 1 summarizes heterogeneity by commitment to one's field of study, measured at program entry by whether students would change their branch or department if choosing again. The corresponding estimates are reported in Supplementary Table 7. Among female students who report that they would change fields, being taught by a female instructor raises course grades by 0.305 SD (95% CI: 0.131 to 0.479 SD; $P < 0.001$). Being taught by a female instructor has no statistically significant effect for students who would choose the same field (0.049 SD; 95% CI: -0.015 to 0.113 SD; $P = 0.14$). The corresponding standardized-test estimates are 0.032 SD (95% CI: -0.001 to 0.065 SD; $P = 0.05$) among students who would choose the same field and 0.025 SD (95% CI: -0.056 to 0.106 SD; $P = 0.54$) among students who would change fields. We also examine heterogeneity by other measures of academic preparation and psychological orientation toward STEM at program entry. The course-grade effects are large and statistically significant for students with lower prior achievement and lower STEM-related confidence, but there is no statistically significant effect for students with higher STEM-related anxiety (Supplementary Table 8).

Table 1: Effect of female faculty on female students' academic achievement.

	I Course grades	II Standardized test score
Female faculty × Female student	0.083***	
	(0.032)	
Female faculty share × Female student		0.033**
		(0.015)
Female faculty share		-0.010
		(0.012)
Female student		0.009
		(0.067)
Student fixed effects	Yes	No
Classroom fixed effects	Yes	No
Department fixed effects	No	Yes
Subject fixed effects	No	Yes
Faculty-level controls	No	Yes
Student-level controls	No	Yes
Observations	28,558	1,752

Notes: This table reports the estimated effects of female faculty on female students' academic outcomes. Column I reports estimates from linear regression estimated by OLS for course grades using Equation (1). The unit of observation is a student–course pair. The specification includes student fixed effects and classroom fixed effects, and standard errors (in parentheses) are clustered at the faculty level. Column II reports estimates from linear regression estimated by OLS for standardized mathematics and science test scores using Equation (3). The unit of observation is a student–subject pair; students are observed once for mathematics and once for science. The specification includes department/school fixed effects, subject fixed effects, student-level controls, average faculty-level controls, and the baseline test score in the corresponding subject. Standard errors are clustered at the student level. Female faculty share is scaled so that one unit equals 10 percentage points. Both outcomes are reported as z-scores (mean = 0, SD = 1). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Figure 1: Effect of female faculty on academic achievement by commitment to field of study. The course-grade panel presents estimates and 95% confidence intervals for the effect of being taught by a female instructor on female students' course-grade z-scores (mean = 0, SD = 1) separately by whether students reported that they would change their branch or department if choosing again. The test-score plot presents estimates and 95% confidence intervals for the effect of a 10-percentage-point increase in the share of courses taught by female faculty over two years on female students' standardized test scores for the same two groups.



Psychological outcomes and beliefs

Our results indicate that greater exposure to female faculty lowers female students' STEM-related anxiety, with no statistically significant effect on STEM-related confidence (Table 2). A 10 percentage point increase in female faculty exposure over two years lowers female students' STEM-related anxiety by 0.053 SD (95% CI: 0.022 to 0.084 SD; $P < 0.001$) relative to male students. The corresponding estimate for STEM-related confidence is 0.001 SD (95% CI: -0.028 to 0.030 SD; $P = 0.95$). This reduction suggests that anxiety can respond to classroom cues that emerge during college, particularly in male-dominated STEM settings. The pattern is consistent with exposure to female faculty functioning as an identity-safety cue that reduces perceived threat in STEM environments.

Beyond anxiety and confidence, we examine effects on classroom engagement and commitment to one's field of study. Supplementary analyses report a 2.5 percentage point change in the gender gap in classroom engagement (95% CI: -0.1 to 5.1 percentage points; $P = 0.06$), which is not statistically significant. The estimated effect on commitment to one's field of study (0.8 percentage points; 95% CI: -1.2 to 2.8 percentage points; $P = 0.43$) is also not statistically significant (Supplementary Table 9).

Table 2: Effect of female faculty on female students' STEM-related anxiety and confidence.

	I STEM-related anxiety	II STEM-related confidence
Female faculty share × Female student	-0.053***	0.001
	(0.016)	(0.015)
Female faculty share	0.015	-0.002
	(0.014)	(0.013)
Female student	0.044	0.010
	(0.077)	(0.068)
Department fixed effects	Yes	Yes
Subject fixed effects	Yes	Yes
Student-level controls	Yes	Yes
Faculty-level controls	Yes	Yes
Observations	1,749	1,749

Notes: This table reports the effects of female faculty on female students' STEM-related anxiety and confidence. Estimates are from linear regression estimated by OLS at the student–subject level, with one observation for mathematics and one for science. Outcomes are endline STEM-related anxiety and confidence indices, measured separately for mathematics and science and reported as z-scores (mean = 0, SD = 1). All specifications include the baseline value of the corresponding outcome, student-level controls, average faculty-level controls, department/school fixed effects, and subject fixed effects. Standard errors (in parentheses) are clustered at the student level. Female faculty share is scaled so that one unit equals 10 percentage points. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Exposure to female faculty also shifts students' gender beliefs about ability in STEM. Table 3 shows that students become more likely to endorse non-stereotypical views, with larger changes among male students, who shift 2.6 percentage points (95% CI: 1.6 to 3.8 percentage points; $P < 0.001$), compared with 1.3 percentage points among female students (95% CI: 0.2 to 2.5 percentage points; $P = 0.04$). Female students' explicit stereotype endorsement is low overall: 85.2 percent endorse non-stereotypical beliefs about women's mathematics ability, compared with 61.1 percent among male students.

We also examine whether male classmates' gender beliefs are associated with female students' anxiety and confidence. Estimating this relationship directly is empirically challenging because male classmates' gender beliefs may be shaped by the same contemporaneous instructors and classroom experiences that also influence female students' anxiety and confidence. We therefore use an instrumental variables design [43], leveraging male classmates' leave-one-classroom-out exposure to female instructors in other first- and second-year courses as a more exogenous source of variation in male peers' stereotype endorsement. The estimated coefficient on male peers' stereotypical beliefs is 0.049 SD for STEM-related anxiety (95% CI: -0.004 to 0.102 SD; $P = 0.07$) and -0.034 SD for STEM-related confidence (95% CI: -0.089 to 0.021 SD; $P = 0.22$; Supplementary Table 10). Neither estimate is statistically significant. We interpret these estimates cautiously because male

peers' exposure to female instructors could also shape female students' outcomes through changes in peers' behavior that are not fully captured by our belief measures. Estimation details are provided in Supplementary Methods.

Table 3: Effect of female faculty on students' beliefs about female ability in STEM.

	Non-stereotypical beliefs
Female faculty share × Female student (β_1)	-0.013**
	(0.006)
Female faculty share (β_2)	0.026***
	(0.005)
Female student (β_3)	0.260***
	(0.029)
$\beta_1 + \beta_2$	0.013**
	(0.006)
P-value ($H_0 : \beta_1 + \beta_2 = 0$)	0.040
Department fixed effects	Yes
Subject fixed effects	Yes
Student-level controls	Yes
Faculty-level controls	Yes
Female students mean	0.852
Male students mean	0.611
Observations	2,581

Notes: This table reports estimates of the effect of female faculty on students' beliefs about female ability in STEM using Equation (3), where non-stereotypical beliefs are coded as 1 and stereotypical beliefs as 0. Beliefs about mathematics were collected from all respondents, and a randomly selected half of the sample were administered the corresponding question about science. Estimates are from linear probability model estimated by OLS. The unit of observation is a student–subject pair (mathematics or science). The specification includes department/school fixed effects, subject fixed effects, student-level controls, and average faculty-level controls. Standard errors (in parentheses) are clustered at the student level. The joint test ($H_0 : \beta_1 + \beta_2 = 0$) assesses whether the total effect of female faculty on the beliefs of female students differs significantly from zero. Female faculty share is scaled so that one unit equals 10 percentage points. The table also reports the mean proportion of non-stereotypical beliefs among female and male students. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Faculty student interactions and teaching practices

To explore potential channels underlying the effects of female faculty on student outcomes, we examine faculty–student interactions and teaching practices. There are no statistically significant

differences between female and male instructors in calling on female students, engaging in informal discussion, or offering office hours, or in directing additional attention toward lower-performing female students (Supplementary Table 11). We also examine self-reported teaching-related activities and practices. For activities with statistically significant gender differences, female faculty report fewer hours than male faculty: female faculty report spending 0.671 fewer hours per week advising students (95% CI: −1.227 to −0.115 hours; $P = 0.02$), 1.087 fewer hours planning lessons (95% CI: −2.067 to −0.108 hours; $P = 0.03$), and 1.482 fewer hours teaching classes (95% CI: −2.814 to −0.150 hours; $P = 0.03$). These differences do not support greater time spent advising, planning lessons, or teaching as an explanation for the effects of female faculty. The only statistically significant difference in Teaching Practices Inventory scores is a 0.229 SD higher score for female faculty on the assignments component (95% CI: 0.023 to 0.435 SD; $P = 0.03$; Supplementary Table 12). Grading bias is unlikely because end-of-semester exams are graded by bodies external to the students' colleges, rather than by the course instructors, which limits opportunities for differential treatment by student gender.

We test whether the female–female advantage varies with instructors' beliefs and observable credentials. Active channels could arise if faculty who hold stereotypical views about gender and ability or fixed mindsets about intelligence systematically allocate time and attention based on inaccurate forecasts about student ability in ways that disadvantage female students [44–46]. However, we find no statistically significant differences in the impact of female faculty on female students between instructors with stereotypical and non-stereotypical beliefs (Supplementary Table 13) or between instructors with fixed and growth mindsets (Supplementary Table 14). Lastly, splitting instructors by academic rank and doctoral degree, we find positive estimates across all categories, with larger point estimates for less senior and non-PhD female faculty (Supplementary Table 15).

3 Discussion

Women tend to earn higher grades than men in STEM [47] but remain underrepresented in the most competitive, high-paying fields. Strong academic performance alone does not guarantee equal career opportunities, as women often face greater scrutiny in hiring, promotions, and evaluations [6, 48]. Their qualifications are judged more critically, requiring stronger academic records to be seen as equally competent [49]. Even a small academic advantage, like the one female faculty provide, may help counter biases favoring male candidates with similar or weaker credentials [3]. In this sense, the boost female faculty give to female students may matter beyond the classroom, influencing access to graduate programs, leadership roles, and longer-run career opportunities.

Higher grades do not fully capture underlying STEM skills. Women also often score lower than men on standardized assessments, a pattern frequently attributed to stereotype threat and related psychological processes that can undermine performance [5]. Our findings suggest that exposure to female faculty helps narrow this gap. Students taught by female faculty show measurable gains in standardized test performance, reducing gender disparities in STEM achievement by 18 percent—a

meaningful improvement, particularly given that it arises in the course of routine student-faculty interactions rather than any additional programmatic intervention or training.

The channels through which faculty gender influences student outcomes can be broadly categorized as active or passive [50]. Active pathways involve direct instructional practices such as mentoring, encouragement, or differential engagement that depend on deliberate instructor behavior. Passive pathways, by contrast, require no intentional action but operate through social cues, such as role modeling, that shape students' beliefs about competence or belonging. When these cues are identity-relevant, exposure to a female faculty member may signal that success is attainable for individuals with shared characteristics.

We examine several active mechanisms through which female faculty could improve student outcomes. We find no statistically significant differences between female and male instructors in measured attention or support for female students. There is also limited scope for differential grading because end-of-semester exams are centrally graded rather than marked by course instructors. Finally, the estimated effects of female faculty on female students do not differ significantly by instructors' beliefs about gender and ability or by fixed versus growth mindset.

The effects on STEM-related anxiety and gender beliefs are consistent with a belief-based channel in which the presence of female faculty shifts students' mental models of who belongs and can succeed in STEM [51], and makes classroom environments feel less shaped by male-dominated norms [52]. Using a simple model of how role models influence student beliefs under uncertainty (see Supplementary Note 1), we treat exposure to female faculty as an identity-relevant signal that students interpret through Bayesian learning [53–55]. Female students in STEM fields, who are often exposed to negative gender stereotypes, may underestimate their ability, feel less efficacious, or feel unwelcome. A female faculty member provides a credible counter-signal, particularly for students with higher uncertainty about their place in STEM. In the model, belief updates are largest when new information is both credible because it comes from someone who shares a salient identity, and novel, in that it contrasts with prior experiences in male-dominated environments.

We find that exposure to female faculty improves course grades in the classes they teach. These within-course gains are interpreted as changes in how students experience and respond to the specific classroom where the signal originates. Research on stereotype threat and belonging uncertainty shows that identity-relevant cues in a classroom, such as who is present and whether one feels numerically isolated or scrutinized, can change anxiety, engagement, and performance without altering underlying competence [5, 9, 56]. In our setting, a female student taught by a woman in a STEM course may feel safer to participate, experience less anxiety, and allocate more effort in that specific course, helping to explain the improvement in grades.

Exposure can also shape beliefs in ways that extend beyond any single classroom experience. To examine this, we estimate impacts on standardized assessments administered two years after exposure, outside the context of any particular course or instructor and without instructor-specific cues. Students with greater cumulative exposure to female faculty perform better on this test as well, indicating that exposure leaves a more durable imprint on students' beliefs or psychological

orientation toward STEM [57]. The estimated average effects on both course grades and standardized tests are statistically significant, although the effect on standardized test is smaller. This pattern is consistent with course grades reflecting both immediate, class-specific responses to having a female instructor and emerging shifts in students' broader views about threat and belonging, whereas the standardized assessment is likely more sensitive to these broader beliefs that students carry with them across courses and exam contexts.

Effects are positive and statistically significant among students who report lower commitment to their field of study at program entry, consistent with belief updating under uncertainty. We find large gains among female students who report that they would change their branch or department if choosing again. These students may be more uncertain about their academic interests, expected returns, or fit, so a credible, identity-relevant cue from a female instructor may have more influence on their beliefs and behaviour. As a result, they show large improvements when exposed to female faculty. We also find statistically significant effects among students with lower prior academic performance or confidence. For standardized test scores, neither subgroup estimate is statistically significant.

These findings reconcile with the seemingly contrasting results of Carrell et al. [32] in the Air Force Academy, where the largest impacts are on high-achieving women. In Carrell et al., major choice is flexible, making persistence in STEM the key outcome, and the largest effects appear for high-achieving women deciding whether to remain in STEM. In our fixed-degree engineering programs, major switching is not common, making course performance the main outcome, and large effects appear for lower-achieving women. Viewed together, the two settings suggest a common mechanism: responses are greatest where prior uncertainty is highest in each context—continuation in STEM in one case, academic performance within a constrained program in the other.

We find that exposure to female faculty reduces STEM-related anxiety but has no statistically significant effect on STEM-related confidence. These outcomes may capture distinct dimensions of students' experiences in STEM fields. Self-efficacy reflects perceived capability to learn and perform in a domain [58], whereas identity safety captures expectations of fair evaluation and belonging without stereotype-based questioning of one's competence [56, 59]. In our empirical analysis, we treat STEM-related anxiety as a proxy for psychological threat (the inverse of identity safety) in STEM domains and STEM-related confidence as a measure of self-efficacy in STEM settings. Both constructs can shape behaviour by altering the expected returns to effort and the psychological cost of engagement, for example when students believe that people like them rarely succeed or that speaking up in class risks unfair judgment.

To relate these constructs to achievement, we conduct a simple mediation decomposition. The estimated indirect effects through anxiety and confidence are not statistically significant (Supplementary Table 16). Because only instructor assignment is randomized and the mediators are measured post-treatment, the assumptions needed for causal mediation are unlikely to hold; we treat this decomposition as descriptive rather than causal [60].

Taken together, the results suggest that identity-relevant cues from female faculty can reduce

anxiety and reshape perceptions of who can succeed in STEM—particularly in male-dominated fields such as engineering [1]. For women, exposure is associated with lower STEM-related anxiety, a pattern consistent with mitigating identity threat. Repeated exposure may also contribute to students' perceptions of fit or belonging, but the commitment item used in our empirical analysis does not isolate those constructs. This perspective aligns with the idea of stereotype inoculation, in which consistent cues from in-group exemplars help students reinterpret what success in STEM looks like [12]. For men, exposure shifts beliefs away from gender stereotypes about STEM ability.

Some limitations warrant mention. First, we focus on short- to medium-term outcomes such as course grades, standardized test scores, and shifts in beliefs and do not observe graduate school enrollment or labor-market trajectories. These outcomes still shed light on longer-run trajectories, because course grades may predict persistence in STEM majors, and prior work shows that women's continuation in STEM fields is especially sensitive to their grades [63, 64]. Effects on dropout and major switching are also not informative in this setting: observed dropout during the study period is rare, and no students in the administrative records are observed switching out of their computer science or electrical engineering programs. Second, although our setting covers a diverse set of engineering colleges with a standardized curriculum, the results may not fully generalize to other higher education systems. Yet, mechanisms related to stereotype threat and role-model exposure have been documented in a range of educational contexts, so the qualitative patterns we document may extend to other settings even if effect sizes differ. Moreover, India accounts for nearly a quarter of all engineering degrees awarded globally [34], and its graduates make up a substantial share of the international STEM workforce, including in the United States. The setting is therefore both policy-relevant in its own right and informative about a large segment of the pipeline into global STEM labor markets.

Although our setting is STEM higher education, the dynamics we document are not unique to universities or to India. Students form beliefs about who succeeds by watching who teaches, who holds authority, and who appears to belong. These cues operate in classrooms, training programs, online courses, and workplaces alike. Our results show that when students repeatedly encounter instructors from groups they don't typically associate with success, particularly in settings where such signals are rare or uneven, it can shift what they believe and how they perform.

4 Methods

We collected student, faculty, and administrative data from a nationally representative random sample of 50 engineering and technology colleges in India. Our sample included faculty and students from broadly defined computer science (CS) and electrical engineering (EE) majors, the two largest disciplines in engineering and technology colleges. In 12 of these 50 colleges, students were randomly assigned to faculty-taught sections within courses ("classrooms") for all first- and second-year courses through a formal, computerized procedure. One of these was an all-women institution and was excluded from the analysis, leaving a final study sample of 11 colleges. Supplementary Fig. 1

summarizes the sample construction. Further details on the population of colleges and sampling procedure are available in Loyalka et al. [65].

Students follow a fixed first- and second-year curriculum. They enroll in required courses each semester, and courses are divided into multiple classrooms to maintain small class sizes. Each classroom represents a separate course section taught by a different faculty member in the same semester. For example, Electrical Engineering 101, Spring 2019, at a given college may have three classrooms: Section A taught by Faculty X, Section B by Faculty Y, and Section C by Faculty Z. The number of classrooms per course ranges from 1 to 15, with a median of 3.

Because students are assigned to classrooms rather than choosing instructors, and faculty do not choose students in their classrooms, there is no self-selection into classrooms based on faculty or peer characteristics. This allows us to estimate the causal effect of being taught by female faculty on student outcomes. In addition, since instructor assignment is random across a student's first- and second-year courses, the proportion of courses taught by female faculty is also determined by chance rather than by student choice, which enables estimation of the impact of cumulative exposure to female faculty during the study period.

We combined administrative records, surveys, and standardized assessments to construct a linked student–faculty dataset. Administrative records provided information on student course enrollments, classroom and instructor assignments, and course grades. Any administrative adjustments, withdrawals, or retakes are reflected in these realized records. Student surveys captured demographic details—including gender, parental education, caste or socioeconomic status, and rural or urban background—along with responses on STEM-related confidence, anxiety, and gender attitudes. To measure academic performance, we linked student records to standardized test scores from a mathematics and science (physics) assessment administered to a randomly selected half of the students at the start of the program and again after two years. We use baseline to refer to measures collected at program entry and endline to refer to measures collected approximately two years later. STEM-related anxiety, confidence and gender attitudes were also measured for this subsample. All students in the sample were enrolled in engineering programs and completed mathematics and science coursework during the first two years of the curriculum. These subjects therefore provide a common academic scale across institutions, departments, and classrooms, measured independently of course instructors and local grading practices. Faculty surveys provided information on rank, qualifications, and demographic characteristics, as well as self-reported teaching activities and instructional practices. These faculty data were linked to the students they taught in the first two years.

In the Indian engineering college system, faculty who teach undergraduate courses hold a range of qualifications, reflecting national hiring norms. While a doctoral degree is common, it is not uniformly required: instructors may hold a master's degree in engineering (M.Tech) and, in some cases, be concurrently enrolled in a Ph.D. program. This practice is consistent with national faculty eligibility criteria and reflects the emphasis on teaching rather than research in undergraduate engineering programs.

The study received institutional support from key government agencies, including the Ministry of Human Resource Development and the All India Council for Technical Education, as well as from college and department administrators. A large team of enumerators was trained to proctor the surveys and assessments in person at each participating college at the start of the first year and the end of the second year. Enumerators were not informed of the study hypotheses, and students and faculty were not told the specific hypotheses being tested. We achieved high participation rates of approximately 95 percent of enrolled students in both survey rounds. The human-subjects data collection was approved by the Stanford University Institutional Review Board (IRB no. 31585).

Estimation strategy

We investigate the impacts of female faculty on student academic and psychological outcomes using a set of regression models. First, we estimate the impact of being randomly assigned to a female instructor's classroom on course grades using Equation (1):

$$Y_{ikcf} = \alpha + \beta_1 \, Fem_f \times Fem_i + \lambda_i + \lambda_c + \epsilon_{ikcf} \tag{1}$$

Here, Y_{ikcf} is the standardized course rank (in SD units) for student i in course k, assigned to classroom c taught by faculty f; Fem_f indicates whether faculty f is female; Fem_i indicates whether student i is female; λ_i are student fixed effects; and λ_c are classroom fixed effects. The coefficient of interest, β_1, estimates whether the female–male difference in standardized course grades within a classroom is more favorable to female students when the instructor is female rather than male. Equivalently, it captures the effect of being taught by a female faculty member rather than a male faculty member on female students' grades relative to male classmates' grades. Student gender is absorbed by student fixed effects, and faculty gender is absorbed by classroom fixed effects. The unit of observation is a student–course, and standard errors are clustered at the faculty level. Faculty gender is not included separately in this specification, as it is subsumed by classroom fixed effects.

We also estimate an auxiliary version of Equation (1) that omits classroom fixed effects, includes course fixed effects, and adds linear controls for faculty characteristics where indicated to examine the effect of female faculty on both male and female students [26]:

$$Y_{ikcf} = \alpha + \beta_1 \, Fem_f \times Fem_i + \beta_2 \, Fem_f + \lambda_i + \lambda_k + \epsilon_{ikcf}. \tag{2}$$

In this specification, λ_k are course fixed effects, and β_1 measures the additional improvement in course grades for female students taught by a female faculty member rather than a male faculty member, relative to the effect on male students.

Equation (3) estimates the effect of being taught by a greater share of female faculty during the first two years on endline test scores, conditional on observable student and faculty characteristics,

including baseline scores:

$$\begin{aligned} Y_{ils} = \alpha &+ \beta_1 FemShare_{is} \times Fem_i \\ &+ \beta_2 FemShare_{is} + \beta_3 Fem_i \\ &+ \gamma^{\intercal}\bar{\Omega}_{is} + \delta^{\intercal}\eta_i + \lambda_l + \lambda_s + \epsilon_{ils}. \end{aligned} \tag{3}$$

Here, Y_{ils} is the standardized test score in subject $l \in \{\text{mathematics, science}\}$ for student i in school/department s. Each student can therefore contribute up to two observations, one for mathematics and one for science. Outcomes are standardized within subject, subject fixed effects are included, and standard errors are clustered at the student level. $FemShare_{is}$ is the proportion of first- and second-year courses in which student i had a female faculty member (scaled by a factor of 10, so that a one-unit increase corresponds to a 10 percentage point increase in exposure), $\bar{\Omega}_{is}$ is a vector of average faculty characteristics for student i in school s, η_i is a vector of student characteristics, λ_l are subject fixed effects, and λ_s are school/department fixed effects.

Beyond academic outcomes, we examine how exposure to female faculty influences students' STEM-related anxiety, confidence, commitment to their field of study, and gender beliefs about ability in mathematics and science. For anxiety, confidence, and gender beliefs, which are elicited separately for mathematics and science, we stack subject-specific indices and estimate the same specification as in Equation (3), with Y_{ils} now denoting a standardized index of the psychological outcome for student i in school/department s and subject $l \in \{\text{mathematics, science}\}$, and with $FemShare_{is}$ scaled as above. These specifications also include subject fixed effects and cluster standard errors at the student level.

For commitment to one's field of study, which is not subject-specific, we again use the specification in Equation (3) without stacking across subjects and without subject fixed effects, so that Y_{is} denotes the outcome for student i in school/department s.

We estimated subgroup effects by baseline ability, commitment to one's field of study at program entry, STEM-related confidence, and STEM-related anxiety using Equation (1) to examine whether the effect of female faculty varies by students' academic standing and psychological orientation toward mathematics and science.

We present two supplementary analyses. First, we use a two-stage least squares specification to relate male peers' gender beliefs to female students' STEM-related anxiety and confidence. The instrument uses male peers' leave-one-classroom-out exposure to female faculty in their other first- and second-year courses, interacted with a mathematics-versus-science contrast. Second, we estimate a descriptive mediation analysis among female students to assess whether STEM-related anxiety or confidence accounts for part of the relationship between female faculty exposure and endline standardized test performance. Details of both specifications are provided in Supplementary Methods.

We addressed missing data in control variables by imputing all missing values with zero and including missingness indicators in all regressions to account for potential differences between re-

spondents and non-respondents. This approach preserves sample size while allowing flexible adjustment for non-random missingness. Our results are robust to excluding these controls or restricting the sample to observations with complete data.

Statistics

Statistical tests of regression coefficients, tests of linear combinations, and Kolmogorov–Smirnov tests are two-sided and use an alpha level of 0.05. Pearson chi-squared tests and omnibus permutation tests use upper-tail P-values. Continuous outcomes are estimated using linear regressions by ordinary least squares (OLS), binary outcomes using linear probability models estimated by OLS, and the supplementary instrumental variable analysis using two-stage least squares. Standard errors are heteroskedasticity-robust or clustered at the level stated in each table note. Table notes report the model, unit of observation, fixed effects, controls, standard error adjustments, and number of observations for each analysis. Estimates discussed in the text are reported with 95% confidence intervals and exact P-values unless $P < 0.001$.

Measurement of outcomes

Course grades in our sample are primarily based on end-of-semester exams graded by a centralized university authority, rather than by individual faculty. This ensures that faculty within the same course do not directly control student assessments. In a few departments, faculty influence a small portion of the grading structure, but this never exceeds 30 percent. Grading systems vary across colleges, with some using letter grades and others numerical scores from 1 to 100. To harmonize grading across institutions, we construct within-course ranks (across all classrooms taught by different faculty in the same course, college, and semester). By construction, these ranks follow an approximately uniform distribution on 0–100 (mean 50, SD 28.9). We then standardize course ranks to have mean zero and standard deviation one for ease of interpretation.

We next estimate the impacts of exposure to female faculty on students' performance on standardized tests of academic skills (mathematics and science), administered by the research team at program entry (baseline) and again after two years (endline).

We use STEM-related anxiety as a proxy for psychological threat (or the inverse of identity safety) in mathematics and science. This index aggregates self-reported worries and nervousness about learning and performing in mathematics and science, which prior research links to affective responses triggered by low identity safety or stereotype threat in academic contexts [61, 62]. We use STEM-related confidence as a proxy for self-efficacy, capturing students' beliefs in their capabilities to learn and succeed in mathematics and science [66]. We measured STEM-related anxiety and confidence using survey items assessing students' emotional responses and self-perceived abilities in mathematics and science. The anxiety index included five items capturing worry and nervousness about learning and performing in the subject, including feeling lost when working on problem sets. The confidence index comprised four items capturing students' perceptions of their own ability (e.g., getting good grades, seeing themselves as fast learners, viewing mathematics or science as a

strong subject, and feeling capable of understanding difficult content). The anxiety items follow the OECD PISA Mathematics Anxiety scale, and the confidence items were adapted from OECD PISA Mathematics Self-Concept items [67, 68], with parallel wording for science (Supplementary Table 17). Participants responded on a 4-point Likert scale; items were standardized and sign-oriented so that higher values indicate more of the construct. Cronbach's α ranged from 0.77 to 0.79 for the confidence scales and from 0.75 to 0.77 for the anxiety scales [69]. Supplementary Table 17 reports the survey items used to construct the STEM-related anxiety and STEM-related confidence indices.

We summarize each multi-item scale with the Anderson inverse-covariance–weighted (ICW) index [40]. Let Z_i denote the vector of standardized item scores for unit i, and let Σ be their variance–covariance matrix. We form the index in Equation (4):

$$I_i \;=\; w^\top Z_i, \qquad w \propto \Sigma^{-1}\mathbf{1}, \qquad \mathbf{1}^\top w = 1, \tag{4}$$

which yields the minimum-variance linear composite for testing mean shifts across a correlated family of items. This construction (i) maximizes signal-to-noise and statistical power, (ii) down-weights redundant or noisy items while up-weighting more precise ones, (iii) accounts for inter-item correlation, and (iv) collapses multiple tests into a single outcome. Equal-weight averages are a special case, so ICW is weakly more efficient whenever items differ in reliability or are correlated. As a robustness check, Supplementary Table 18 shows that results are nearly identical when using principal component analysis to construct the indices.

Commitment to one's field of study is measured using the inverse of the question: *Would you change your branch/department if you could choose again?* The measure equals one if a student reports that they would choose the same field and zero if they would change fields. The single item may reflect academic interests, comparative advantage, curriculum satisfaction, career plans, expected returns, perceived fit, belonging, or a combination of these considerations [71]. Classroom engagement was measured based on how often students reported asking questions in class. Students were classified as engaged if they participated at least once every few classes in both mathematics and science classes.

Student beliefs about gender differences in ability in mathematics and science were recorded using a five-point Likert scale. Responses ranged from 1 (strong belief that male students are much better) to 5 (strong belief that female students are much better). These responses were recoded into a binary variable, classifying beliefs as stereotypical if students expressed that male students are better or much better, and non-stereotypical otherwise.

Faculty beliefs were measured similarly, eliciting their perceptions regarding student qualifications in mathematics and science at program entry and after two years, as well as student performance in major-specific core courses (apart from mathematics and science) after two years.

Faculty mindset was measured using two constructs. Fixed mindset was assessed with the statement *Intelligence is something that cannot be changed very much.* Faculty who strongly agreed or agreed were coded as having a fixed mindset, while the rest were classified as not having a fixed

mindset (having a growth mindset). A second item captured a related belief that differentiates between learning skills and changing intelligence, using the statement *You can learn new things, but you can't change a person's intelligence*. Faculty who strongly agreed or agreed were again coded as having a fixed mindset, while the rest were coded as not having a fixed mindset.

Faculty surveys provided information on rank, qualifications, demographic characteristics, time spent on teaching-related activities, and self-reported instructional practices, including items from the Teaching Practices Inventory [70]. Student surveys measured classroom experiences and faculty-student interactions, including whether students were called on in class, discussed coursework during breaks, and attended office hours. These survey measures were linked to administrative records on course enrollments, instructor assignments, and grades for first- and second-year courses. Supplementary Table 19 summarizes the timing and data sources of the main variables used in the analysis.

Data Availability

The data used to perform the analyses will be deposited in Open Science Framework.

Code Availability

The code used to perform the analyses will be deposited in Open Science Framework.

References

[1] Sapna Cheryan, Sianna A. Ziegler, Amanda K. Montoya, and Lily Jiang. Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, 143(1):1, 2017.

[2] Kim Parker. What's behind the growing gap between men and women in college completion?, November 8, 2021. URL `https://www.pewresearch.org/short-reads/2021/11/08/whats-behind-the-growing-gap-between-men-and-women-in-college-completion/`.

[3] Marianne Bertrand, Claudia Goldin, and Lawrence F. Katz. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–255, 2010.

[4] National Center for Science and Engineering Statistics. Diversity and STEM: Women, Minorities, and Persons with Disabilities 2023. URL `https://ncses.nsf.gov/pubs/nsf23315/report`. Accessed: 2024-08-08.

[5] Claude M. Steele and Joshua Aronson. Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5):797, 1995.

[6] Ernesto Reuben, Paola Sapienza, and Luigi Zingales. How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408, 2014. ISSN 0027-8424. doi: 10.1073/pnas.1314788111. URL `https://www.pnas.org/content/111/12/4403`.

[7] Muriel Niederle and Lise Vesterlund. Gender and competition. *Annual Review of Economics*, 3(1):601–630, 2011.

[8] Pedro Bordalo, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. Beliefs about gender. *American Economic Review*, 109(3):739–773, March 2019. doi: 10.1257/aer.20170007. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20170007`.

[9] Mary C. Murphy, Claude M. Steele, and James J. Gross. Signaling threat: how situational cues affect women in math, science, and engineering settings. *Psychological Science*, 18(10):879–885, 2007.

[10] Beth Mitchneck, Jessi L. Smith, and Melissa Latimer. A recipe for change: creating a more inclusive academy. *Science*, 352(6282):148–149, 2016.

[11] Marianne Bertrand and Esther Duflo. Field experiments on discrimination. *Handbook of Economic Field Experiments*, 1:309–393, 2017.

[12] Nilanjana Dasgupta. Ingroup experts and peers as social vaccines who inoculate the self-concept: the stereotype inoculation model. *Psychological Inquiry*, 22(4):231–246, 2011.

[13] Jane G. Stout, Nilanjana Dasgupta, Matthew Hunsinger, and Melissa A. McManus. STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, 100(2):255–270, 2011. ISSN 1939-1315, 0022-3514. doi: 10.1037/a0021385. URL https://doi.apa.org/doi/10.1037/a0021385.

[14] Nilanjana Dasgupta, Melissa McManus Scircle, and Matthew Hunsinger. Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences*, 112(16):4988–4993, April 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1422822112. URL https://pnas.org/doi/full/10.1073/pnas.1422822112.

[15] Deborah J. Wu, Kelsey C. Thiem, and Nilanjana Dasgupta. Female peer mentors early in college have lasting positive impacts on female engineering students that persist beyond graduation. *Nature Communications*, 13(1):6837, November 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34508-x. URL https://www.nature.com/articles/s41467-022-34508-x.

[16] Sarah D. Herrmann, Robert Mark Adelman, Jessica E. Bodford, Oliver Graudejus, Morris A. Okun, and Virginia S. Y. Kwan. The effects of a female role model on academic performance and persistence of women in STEM courses. *Basic and Applied Social Psychology*, 38(5):258–268, September 2016. ISSN 0197-3533, 1532-4834. doi: 10.1080/01973533.2016.1209757. URL https://www.tandfonline.com/doi/full/10.1080/01973533.2016.1209757.

[17] Penelope Lockwood. "Someone like me can be successful": do college students need same-gender role models? *Psychology of Women Quarterly*, 30(1):36–46, 2006.

[18] Robert W. Fairlie, Florian Hoffmann, and Philip Oreopoulos. A community college instructor like me: race and ethnicity interactions in the classroom. *American Economic Review*, 104(8):2567–2591, 2014. URL https://doi.org/10.1257/aer.104.8.2567.

[19] Serena Canaan and Pierre Mouganie. The impact of advisor gender on female students' STEM enrollment and persistence. *Journal of Human Resources*, 58(2):593–632, 2023.

[20] Eric P. Bettinger and Bridget Terry Long. Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review*, 95(2):152–157, 2005. URL https://doi.org/10.1257/000282805774670149.

[21] Florian Hoffmann and Philip Oreopoulos. A professor like me: the influence of instructor gender on college achievement. *Journal of Human Resources*, 44(2):479–494, 2009. ISSN 0022-166X. doi: 10.3368/jhr.44.2.479.

[22] Bruna Borges and Fernanda Estevan. Does exposure to more women in male-dominated fields render female students more career-oriented? *Management Science*, 71(6):5045–5082, 2025. doi: 10.1287/mnsc.2023.01422.

[23] Thomas S. Dee. Teachers and the gender gaps in student achievement. *Journal of Human Resources*, 42(3):528–554, 2007. doi: 10.3368/jhr.XLII.3.528.

[24] Jaegeum Lim and Jonathan Meer. The impact of teacher–student gender matches: random assignment evidence from South Korea. *Journal of Human Resources*, 52(4):979–997, 2017. doi: 10.3368/jhr.52.4.1215-7585R1.

[25] Jaegeum Lim and Jonathan Meer. Persistent effects of teacher–student gender matches. *Journal of Human Resources*, 55(3):809–835, 2020. doi: 10.3368/jhr.55.3.0218-9314R4.

[26] Karthik Muralidharan and Ketki Sheth. Bridging education gender gaps in developing countries: the role of female teachers. *Journal of Human Resources*, 51(2):269–297, 2016. URL `https://doi.org/10.3368/jhr.51.2.0813-5901R1`.

[27] Jie Gong, Yi Lu, and Hong Song. The effect of teacher gender on students' academic and noncognitive outcomes. *Journal of Labor Economics*, 36(3):743–778, 2018. doi: 10.1086/696203.

[28] David Card, Ciprian Domnisoru, Seth G. Sanders, Lowell Taylor, and Sharyn J. Udalova. The impact of female teachers on female students' lifetime well-being. Working Paper 30430, National Bureau of Economic Research, 2022.

[29] Michela Carlana. Implicit stereotypes: evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3):1163–1224, 2019. ISSN 0033-5533. doi: 10.1093/qje/qjz008. URL `https://doi.org/10.1093/qje/qjz008`.

[30] Victor Lavy and Edith Sand. On the origins of gender gaps in human capital: short- and long-term consequences of teachers' biases. *Journal of Public Economics*, 167:263–279, 2018. doi: 10.1016/j.jpubeco.2018.09.007.

[31] Sonali Rakshit and Soham Sahoo. Biased teachers and gender gap in learning outcomes: evidence from India. *Journal of Development Economics*, 161:103041, 2023. doi: 10.1016/j.jdeveco.2022.103041.

[32] Scott E. Carrell, Marianne E. Page, and James E. West. Sex and science: how professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125(3):1101–1144, 2010. ISSN 0033-5533. URL `https://doi.org/10.1162/qjec.2010.125.3.1101`.

[33] Hani Mansour, Daniel I. Rees, Bryson M. Rintala, and Nathan N. Wozny. The effects of professor gender on the postgraduation outcomes of female students. *ILR Review*, 75(3):693–715, 2022.

[34] National Science Board. *Science and Engineering Indicators*. 2016.

[35] Ministry of Education, Government of India. All India Survey on Higher Education 2021–22. Technical report, Department of Higher Education, Ministry of Education, Government of India, New Delhi, India, 2024. URL https://aishe.gov.in/aishe-final-report/.

[36] National Science Board, National Science Foundation. Higher Education in Science and Engineering. Technical Report NSB-2023-32, National Center for Science and Engineering Statistics, National Science Foundation, Alexandria, VA, 2023. URL https://ncses.nsf.gov/pubs/nsb202332.

[37] National Statistical Office, Government of India. Periodic Labour Force Survey: Annual Report, July 2023–June 2024. Technical report, Ministry of Statistics and Programme Implementation, Government of India, New Delhi, India, 2024. URL https://www.mospi.gov.in/sites/default/files/publication_reports/AnnualReport_PLFS2023-24L2.pdf.

[38] Department of Science and Technology, Government of India. Research and Development Statistics at a Glance 2022–23. Technical report, Ministry of Science and Technology, Government of India, New Delhi, India, 2023. URL https://dst.gov.in/document/reports/updated-rd-statistics-glance-2022-23.

[39] Prashant Loyalka, Ou Lydia Liu, Guirong Li, Elena Kardanova, Igor Chirikov, Shangfeng Hu, Ningning Yu, Liping Ma, Fei Guo, Tara Beteille, et al. Skill levels and gains in university STEM education in China, India, Russia and the United States. *Nature Human Behaviour*, 5 (7):892–904, 2021. URL https://doi.org/10.1038/s41562-021-01062-3.

[40] Michael L. Anderson. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495, 2008. doi: 10.1198/016214508000000841. URL https://doi.org/10.1198/016214508000000841.

[41] Rachel Glennerster and Kudzai Takavarasha. *Running Randomized Evaluations: A Practical Guide*. Princeton University Press, 2014.

[42] Ralph Stinebrickner and Todd R. Stinebrickner. The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis & Policy*, 8(1), 2008.

[43] Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

[44] Sarah-Jane Leslie, Andrei Cimpian, Meredith Meyer, and Edward Freeland. Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219): 262–265, 2015. doi: 10.1126/science.1261375. URL https://www.science.org/doi/abs/10.1126/science.1261375.

[45] David S. Yeager, Paul Hanselman, Gregory M. Walton, Jared S. Murray, Robert Crosnoe, Chandra Muller, Elizabeth Tipton, Barbara Schneider, Chris S. Hulleman, Cintia P. Hinojosa, et al. A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774):364–369, 2019.

[46] Robert Rosenthal and Lenore Jacobson. Pygmalion in the classroom. *The Urban Review*, 3 (1):16–20, 1968.

[47] Daniel Voyer and Susan D. Voyer. Gender differences in scholastic achievement: a meta-analysis. *Psychological Bulletin*, 140(4):1174, 2014.

[48] Corinne A. Moss-Racusin, John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41):16474–16479, 2012.

[49] J. Aislinn Bohren, Alex Imas, and Michael Rosenberg. The dynamics of discrimination: theory and evidence. *American Economic Review*, 109(10):3395–3436, 2019.

[50] Thomas S. Dee. A teacher like me: does race, ethnicity, or gender matter? *American Economic Review*, 95(2):158–165, 2005. URL `https://doi.org/10.1257/000282805774670446`.

[51] Sapna Cheryan, Victoria C. Plaut, Paul G. Davies, and Claude M. Steele. Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, 97(6):1045, 2009.

[52] Sapna Cheryan and Hazel Rose Markus. Masculine defaults: identifying and mitigating hidden cultural biases. *Psychological Review*, 127(6):1022, 2020.

[53] Roland Bénabou and Jean Tirole. Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915, 2002.

[54] Erik Eyster and Matthew Rabin. Naive herding in rich-information settings. *American Economic Journal: Microeconomics*, 2(4):221–243, 2010.

[55] Alex Eble and Feng Hu. Child beliefs, societal beliefs, and teacher-student identity match. *Economics of Education Review*, 77:101994, 2020.

[56] Gregory M. Walton and Geoffrey L. Cohen. A question of belonging: race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1):82, 2007.

[57] David S. Yeager and Gregory M. Walton. Social-psychological interventions in education: they're not magic. *Review of Educational Research*, 81(2):267–301, 2011.

[58] Albert Bandura. Self-efficacy mechanism in human agency. *American Psychologist*, 37(2):122, 1982.

[59] Geoffrey L. Cohen and Julio Garcia. Identity, belonging, and achievement: a model, interventions, implications. *Current Directions in Psychological Science*, 17(6):365–369, 2008.

[60] Julia M. Rohrer, Paul Hünermund, Ruben C. Arslan, and Malte Elson. That's a lot to process! Pitfalls of popular path models. *Advances in Methods and Practices in Psychological Science*, 5(2):25152459221095827, 2022.

[61] Paul G. Davies, Steven J. Spencer, and Claude M. Steele. Clearing the air: identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, 88(2):276, 2005.

[62] Jason W. Osborne. Linking stereotype threat and anxiety. *Educational Psychology*, 27(1): 135–154, 2007.

[63] Ben Ost. The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*, 29(6):923–934, 2010.

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

[65] Prashant Loyalka, Zhaolei Shi, Guirong Li, Elena Kardanova, Igor Chirikov, Ningning Yu, Shangfeng Hu, Huan Wang, Liping Ma, Fei Guo, Ou Lydia Liu, Ashutosh Bhuradia, Saurabh Khanna, Yanyan Li, and Adam Murray. The effect of faculty research on student learning in college. *Educational Researcher*, 51(4):265–273, 2022. ISSN 0013-189X. URL `https://doi.org/10.3102/0013189X221090229`.

[66] Frank Pajares. Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66 (4):543–578, 1996.

[67] Organisation for Economic Co-operation and Development. *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing, Paris, 2023. doi: 10.1787/53f23881-en.

[68] Organisation for Economic Co-operation and Development. PISA 2012 student questionnaire (English–USA). `https://www.oecd.org/content/dam/oecd/en/data/datasets/pisa/pisa-2012-datasets/main-survey/questionnaires/Student%20Questionnaire%20Form%20B%20PISA%202012.pdf`, 2012.

[69] Lee J. Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3): 297–334, 1951.

[70] Carl Wieman and Sarah Gilbert. The teaching practices inventory: a new tool for characterizing college and university teaching in mathematics and science. *CBE–Life Sciences Education*, 13(3):552–569, 2014. `https://doi.org/10.1187/cbe.14-02-0023`.

[71] Anna Womack, Melanie E. Leuty, Emily Bullock-Yowell, and Jon T. Mandracchia. Understanding commitment: relations between major commitment, satisfaction, involvement, fit, university commitment, and intention to quit. *Journal of Career Development*, 45(2):166–182, 2018. doi: 10.1177/0894845316676904.

Acknowledgements

We are grateful to Kathy Baylis, Rajdev Brar, Geoff Cohen, Gabriel Cruz, Ben Domingue, Anmol Gupta, Jalnidh Kaur, Jack Marshall, Paola Sapienza, Olga Stoddard, and Claude Steele for helpful suggestions. We thank seminar participants at UCLA, USC, UCSC, Santa Clara University, the WELAC-RIDGE Gender and Household Economics Workshop, the NSF-SIARM Workshop at the University of Chicago, the All UC Demography Conference, and the Society of Labor Economists Meetings for helpful comments. P.L. acknowledges funding support from the All India Council for Technical Education.

Author Contributions

R.F. and P.L. conceptualized the study; P.L. and S.K. collected survey data. S.K. conducted the analysis of the survey data with inputs from R.F., M.J., P.L., and G.S.; R.F., M.J., P.L., and G.S. wrote the paper. R.F., M.J., P.L., and G.S. provided intellectual input and edited the manuscript. Author names are ordered alphabetically.

Competing Interests

The authors have no competing interests to declare.

Supplementary Information

Female Faculty Reduce Anxiety About STEM and Raise Academic Achievement Among Female Undergraduates

Robert Fairlie, Mridul Joshi, Saurabh Khanna, Prashant Loyalka, and Gagandeep Sachdeva

Supplementary Methods, Supplementary Figures, Supplementary Tables, and Supplementary References

Supplementary Methods

Permutation test for random assignment. We implement a design-based permutation test using the realized student-classroom rosters. The test follows the randomization-test logic in Lehmann and Romano [1] and Good [2], and the classroom-assignment checks used by Carrell and West [3] and Lim and Meer [4]. The null is that, conditional on each course offering and the realized number of students in each classroom, students are exchangeable across classrooms. We define a course offering as a department-course combination, and each such offering corresponds to a unique institution and term. We generate 1,000 placebo assignments by randomly reallocating students across classrooms within each course while preserving realized classroom sizes. For each placebo assignment, we recompute the relevant imbalance statistics and compare the observed statistics to the corresponding randomization distributions.

We report two sets of tests. For continuous baseline covariates, we follow Lim and Meer by computing a classroom-level empirical P-value. For each classroom and covariate, we locate the observed classroom mean in the distribution of placebo classroom means generated under random assignment, using a mid-P adjustment for ties. Under random assignment, these classroom-level P-values should be approximately uniform. We therefore report a Kolmogorov–Smirnov test against the uniform distribution and a Pearson χ^2 test based on ten equal-width bins.

For binary covariates, classroom means are discrete and many ties arise, so the distribution of classroom-level P-values is less informative. We therefore report omnibus permutation tests based on across-classroom dispersion in classroom means. The first statistic sums classroom-size-weighted squared deviations of classroom means from the stratum mean across strata. The second divides each stratum's between-classroom dispersion by its within-stratum variance before summing across strata. The reported permutation P-values are the shares of placebo statistics at least as large as the observed statistic, with a +1 finite-simulation adjustment. Supplementary Table 3 reports the permutation tests, and Supplementary Fig. 2 plots the empirical CDFs of classroom-level P-values for the continuous baseline covariates.

Impact of male peers' gender beliefs. To estimate the relationship between male peers' stereotypical beliefs and female students' psychological outcomes, we construct a student–subject measure W_{id}, for $d \in \{\text{mathematics}, \text{science}\}$, equal to the section-weighted share of male classmates who endorse the gender stereotype in subject d. Let J_i be the set of classrooms where student i is observed, and let M_{jd} be the set of male classmates in section j with observed belief data for subject d. Let N_{jd} denote the number of such male classmates. We define the peer-belief measure

in Supplementary Equation (1):

$$
\begin{aligned}
W_{id} &= \sum_{j \in J_i} \omega_{ijd} \bar{B}_{jd}, \\
\bar{B}_{jd} &= \frac{1}{N_{jd}} \sum_{m \in M_{jd}} B_{md}, \\
\omega_{ijd} &= \frac{N_{jd}}{\sum_{r \in J_i} N_{rd}}.
\end{aligned}
\tag{1}
$$

Here, $B_{md} = 1$ if male classmate m endorses the stereotype in subject d. These weights give more weight to classrooms in which peer beliefs are measured for more male classmates.

As an instrument, we use male peers' leave-one-classroom-out exposure to female faculty in their other observed first- and second-year classrooms. For each male classmate m, let J_m be his set of observed first- and second-year classrooms, and let $F_l = 1$ if classroom l was taught by a female faculty member. For each classroom j that student i shares with male classmate m, we define leave-one-classroom-out exposure in Supplementary Equation (2):

$$
L_{m,-j} = \frac{1}{|J_m| - 1} \sum_{l \in J_m \setminus \{j\}} F_l. \tag{2}
$$

This measure is defined for students with at least two observed classrooms. It excludes the classroom shared with focal student i, so the instrument is not mechanically constructed from the same classroom used to measure the focal student's peer-belief environment.

We then aggregate this leave-one-classroom-out exposure across male classmates. Let M_j^Z be the set of male classmates in section j with nonmissing $L_{m,-j}$, and let N_j^Z be the number of such male classmates. Supplementary Equation (3) defines the resulting instrument:

$$
\begin{aligned}
\bar{L}_j &= \frac{1}{N_j^Z} \sum_{m \in M_j^Z} L_{m,-j}, \\
\nu_{ij} &= \frac{N_j^Z}{\sum_{r \in J_i} N_r^Z}, \\
Z_i &= \sum_{j \in J_i} \nu_{ij} \bar{L}_j.
\end{aligned}
\tag{3}
$$

The resulting instrument Z_i is measured at the student level and does not vary between the mathematics and science rows for a given focal female student.

We estimate the IV model in a stacked two-row panel, with one observation for mathematics and one for science, using Supplementary Equation (4):

$$
Y_{id} = \beta W_{id} + \theta Y_{id}^0 + \alpha_i + \delta_d + \varepsilon_{id}. \tag{4}
$$

Here, Y_{id} is the endline subject-specific outcome, Y_{id}^0 is the corresponding baseline outcome, α_i are student fixed effects, and δ_d are subject fixed effects. The endogenous regressor W_{id} is the share of

male peers endorsing the stereotype in subject d.

Because Z_i is constant across the two subject rows for a given student, it would be absorbed by the student fixed effects if entered directly. We therefore interact Z_i with a mathematics-versus-science contrast D_d, where $D_d = 1$ for mathematics and $D_d = -1$ for science. The first stage is Supplementary Equation (5):

$$W_{id} = \pi(Z_i \times D_d) + \theta_1 Y_{id}^0 + \alpha_i + \delta_d + u_{id}. \tag{5}$$

Identification comes from whether male peers' exposure to female faculty in their other courses predicts differences in their stereotype endorsement between mathematics and science, and whether those predicted differences are associated with female students' corresponding differences in anxiety or confidence.

The exclusion restriction requires that, conditional on student fixed effects, subject fixed effects, and the baseline subject-specific outcome, male peers' leave-one-classroom-out exposure to female faculty affects female students' subject-specific anxiety or confidence only through male peers' measured subject-specific stereotype endorsement. This assumption would be violated if male peers' exposure to female faculty affected female students through peer behavior, informal academic support, classroom participation, or broader peer norms not captured by measured beliefs. Standard errors are clustered at the student level.

Descriptive mediation analysis. We estimate a mediation model among female students to assess whether STEM-related anxiety or confidence statistically account for part of the association between exposure to female faculty and endline academic performance [5, 6]. The treatment is the share of female faculty in a student's prior two years (scaled 0–10), and the mediators are standardized endline measures of anxiety and confidence. The outcome is a standardized score on the endline standardized assessment (mathematics or science). Both mediator and outcome models include controls for student demographics, baseline test scores, baseline levels of the mediator variable, and average faculty characteristics. Standard errors are clustered at the student level. Because the mediators are likely post-treatment and the sequential ignorability assumption is unlikely to hold, we interpret all estimates descriptively as a decomposition of model-implied associations rather than as causal mediation effects.

Computational workflow (pseudocode). The main analysis proceeds as follows. First, read the linked student–course records, harmonize missing-value codes, and construct student, faculty, course, classroom, department, and subject identifiers. Second, construct within-course ranks and standardized course-grade outcomes; standardized mathematics and science test scores; STEM-related anxiety and confidence indices; commitment, classroom-engagement, and gender-belief measures; faculty characteristics, beliefs, mindsets, and teaching-practice measures; female-faculty exposure; and the interactions used in the estimating equations. Third, organize the data at the student–course, student–subject, student, or faculty–department level required for each outcome.

Fourth, estimate the course-grade, standardized-test, anxiety, confidence, commitment, engagement, gender-belief, faculty-interaction, and teaching-practice models using the fixed effects, controls, and standard-error adjustments described in Methods and the table notes. Fifth, estimate the heterogeneity and robustness specifications, the descriptive mediation models, and the two-stage least squares peer-belief models described in Supplementary Methods.

The permutation analysis proceeds separately. Define each randomization stratum by course, institution, and term and retain strata containing multiple classrooms. Within each stratum, repeatedly reassign students across classrooms without replacement while preserving realized classroom sizes. For each continuous baseline covariate, compare the observed classroom mean with its permutation distribution to obtain classroom-level empirical P-values, and test whether those P-values follow a uniform distribution. For each binary baseline covariate, compare the observed across-classroom dispersion with the permutation distribution using the size-weighted and variance-standardized omnibus statistics. Repeat the reassignment 1,000 times and report the resulting uniformity and omnibus permutation tests.

Supplementary Note 1: belief formation model

A female student i takes T_i courses indexed by $c \in \{1, \ldots, T_i\}$. Let $X_{ic} \in \{0, 1\}$ indicate whether student i in course c is taught by a female instructor, and let $\bar{X}_i \equiv T_i^{-1} \sum_c X_{ic} \in [0, 1]$ be the share of courses taught by female instructors up to the time of measurement.

Student performance depends on two psychological state variables: self-efficacy B_i and identity safety S_i. Let $B_i^0 \sim \mathcal{N}(\mu_{B,i}^0, \tau_{B,i}^2)$ and $S_i^0 \sim \mathcal{N}(\mu_{S,i}^0, \tau_{S,i}^2)$ be baseline levels of these states.

When the student is exposed to a female instructor ($X_{ic} = 1$), course c emits two identity-relevant signals—ability (s_{ic}^B) and fit/safety (s_{ic}^S), as shown in Supplementary Equation (6):

$$s_{ic}^B \sim \mathcal{N}(\mu_B, \sigma_B^2), \qquad s_{ic}^S \sim \mathcal{N}(\mu_S, \sigma_S^2). \tag{6}$$

Posterior levels after exposure to s_{ic}^B and s_{ic}^S are updated by the simple learning rule in Supplementary Equation (7):

$$\begin{aligned} B_i^{\text{post}} &= (1 - \kappa_{B,i})\, B_i^{\text{pre}} + \kappa_{B,i}\, s_{ic}^B, \\ S_i^{\text{post}} &= (1 - \kappa_{S,i})\, S_i^{\text{pre}} + \kappa_{S,i}\, s_{ic}^S, \end{aligned} \tag{7}$$

with $\kappa_{B,i} = \frac{\tau_{B,i}^2}{\tau_{B,i}^2 + \sigma_B^2}$ and $\kappa_{S,i} = \frac{\tau_{S,i}^2}{\tau_{S,i}^2 + \sigma_S^2} \in (0, 1)$.

The expected per-exposure update magnitudes $\mathbb{E}[|B_i^{\text{post}} - B_i^{\text{pre}}|]$ and $\mathbb{E}[|S_i^{\text{post}} - S_i^{\text{pre}}|]$ increase with prior uncertainty (larger τ^2), and with distance from prior beliefs, measured by $|\mu_B - \mu_{B,i}^0|$ and $|\mu_S - \mu_{S,i}^0|$.

Let $K_i = \sum_c X_{ic}$ be the number of exposed courses (so $K_i = T_i \bar{X}_i$). Aggregating and approximating

expected per-exposure updates over courses gives Supplementary Equation (8):

$$B_i = B_i^0 + \phi_B(U_i)\,\bar{X}_i, \qquad S_i = S_i^0 + \phi_S(U_i)\,\bar{X}_i, \tag{8}$$

where U_i is a scalar index of baseline uncertainty, and $\phi_B(U_i) \propto T_i\,\kappa_{B,i}\,|\mu_B - \mu_{B,i}^0|$ and $\phi_S(U_i) \propto T_i\,\kappa_{S,i}\,|\mu_S - \mu_{S,i}^0|$ are weakly increasing in U_i. In words, repeated credible cues build the durable psychological states B_i and S_i, and students with higher uncertainty update more with each exposure.

An implication of this framework is that female students may feel greater psychological threat to their STEM identity when exposed to stereotypical peers; correspondingly, S_i can also incorporate peer beliefs. Because our focus is faculty exposure, we do not model this further, as it does not alter the core predictions.

Students are evaluated in two contexts: a course grade in each course c and a standardized test at the end of the period. Course grades include a contemporaneous context-specific effect $\Lambda(U_i)\,X_{ic}$, which operates only in the course setting due to mental association with the instructor's identity, as shown in Supplementary Equation (9):

$$Y_{ic} = \alpha_c + \beta_c\,B_i + \theta_c\,S_i + \Lambda(U_i)\,X_{ic} + \varepsilon_{ic}. \tag{9}$$

The standardized test t, by design, omits this contextual component and loads only the durable stocks, as shown in Supplementary Equation (10):

$$Y_{it} = \alpha_t + \beta_t\,B_i + \theta_t\,S_i + \varepsilon_{it}. \tag{10}$$

The parameters satisfy $\beta_c, \beta_t, \theta_c, \theta_t \geq 0$, and ε_{ic} and ε_{it} are idiosyncratic terms.

The model has the following testable implications:

I1. Course outcomes load both context-specific and durable components, whereas the standardized test loads durable components only. Exposure to female instructors raises course grades via cumulative belief revision and a context-specific term, and raises test scores via cumulative belief revision only, as shown in Supplementary Equation (11):

$$\begin{aligned} \Delta Y_{\text{course}} &= \beta_c\,\phi_B(U_i) + \theta_c\,\phi_S(U_i) + \Lambda(U_i), \\ \Delta Y_{\text{test}} &= \beta_t\,\phi_B(U_i) + \theta_t\,\phi_S(U_i). \end{aligned} \tag{11}$$

Because ΔY_{test} is defined per unit change in exposure *share*, the implied persistent effect per exposed course is $\Delta Y_{\text{test}}/T_i$. Under linear accumulation, a nonnegative contextual term ($\Lambda(U_i) \geq 0$), and weak comparability of stock mappings across contexts (i.e., $\beta_c\phi_B + \theta_c\phi_S \;\geq\; \beta_t\phi_B + \theta_t\phi_S$), Supplementary Equation (12) gives the implication:

$$\Delta Y_{\text{course}} \;\geq\; \frac{\Delta Y_{\text{test}}}{T_i}. \tag{12}$$

Since $T_i \geq 1$, an empirically conservative check is the stricter inequality $\Delta Y_{\text{course}} \geq \Delta Y_{\text{test}}$; if this holds, the per-course prediction is automatically satisfied.

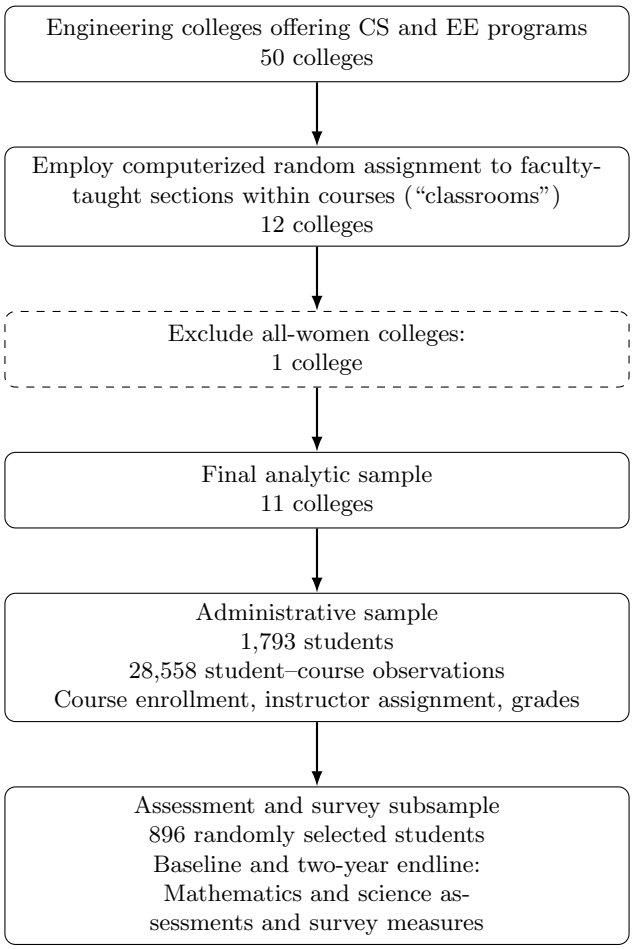
I2. A positive course effect with a null test effect indicates context-specific responses, whereas a positive test effect indicates durable belief change. If $\Delta Y_{\text{test}} > 0$ then at least one durable stock rose (either $\phi_B(U_i) > 0$ or $\phi_S(U_i) > 0$); if $\Delta Y_{\text{course}} > 0 = \Delta Y_{\text{test}}$, the impact of exposure is largely context-specific.

I3. High-uncertainty subgroups should see larger gains on course outcomes. Since $\phi_B(\cdot)$ and $\phi_S(\cdot)$ are weakly increasing in U_i and $\Lambda'(\cdot) \geq 0$—that is, students with more uncertainty update more and receive greater context-specific boosts—the marginal course effect in Supplementary Equation (13) is weakly increasing in U_i:

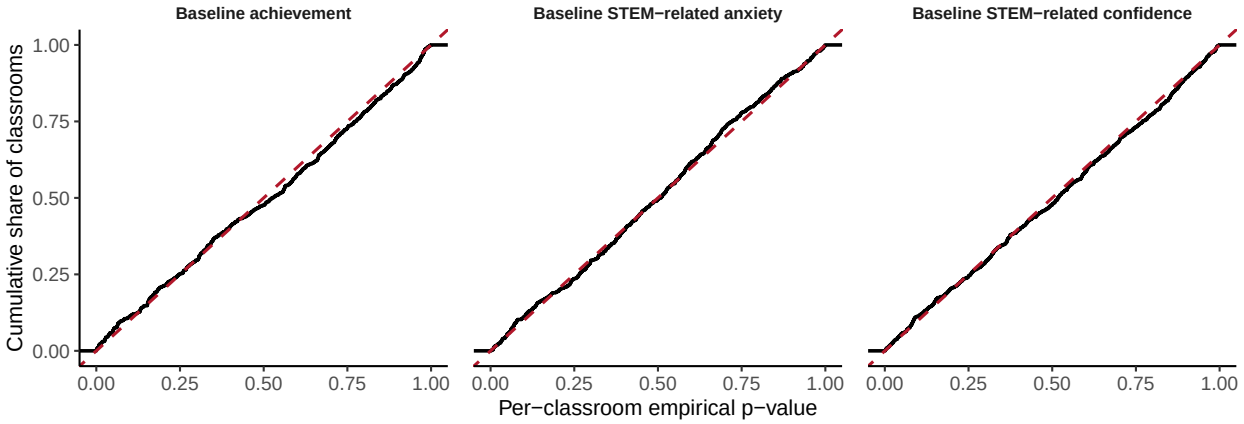
$$\frac{\partial \Delta Y_{\text{course}}}{\partial U_i} = \beta_c \, \phi_B'(U_i) + \theta_c \, \phi_S'(U_i) + \Lambda'(U_i) \;\geq\; 0. \tag{13}$$

The same qualitative pattern should hold on standardized tests, but subgroup gaps are expected to be smaller because (i) the context-specific term drops out and (ii) repeated credible exposure compresses differences in psychological states as priors are pulled toward common reference points.

Supplementary Fig. 1: Sample construction. Flow diagram shows the construction of the analytic sample from the 50 engineering and technology colleges offering computer science (CS) or electrical engineering (EE) programs in the sampling frame. Twelve colleges used computerized random assignment to classrooms; one all-women college is excluded because the empirical strategy compares female and male students within classrooms. The final analytic sample contains 11 colleges with administrative records linking course enrollment, instructor assignment, and grades. The assessment/survey subsample is a randomly selected subset of first-year students from the same colleges and cohorts and is linked to administrative course records.



Supplementary Fig. 2: Distribution of permutation P-values from classroom-level randomization tests. The solid line in each panel shows the empirical cumulative distribution function of classroom-level permutation P-values for the corresponding baseline covariate. The dashed line shows the uniform distribution expected under random assignment. Placebo assignments are generated within course offerings (department-course combinations) by reallocating students across classrooms while preserving realized classroom sizes. Each test uses 1,000 placebo assignments. None of the uniformity tests rejects its null hypothesis at the 5% level.



Supplementary Table 1: Student characteristics and survey measures by gender.

Student characteristic	I Female student	II Male student	III Difference	IV Observations
Mathematics and science proficiency	-0.146	0.042	-0.188***	826
	[0.745]	[1.029]	(0.063)	
STEM-related anxiety	0.053	-0.008	0.061	822
	[0.931]	[1.045]	(0.072)	
STEM-related confidence	-0.057	-0.007	-0.049	822
	[0.912]	[1.063]	(0.072)	
Father attended college	0.554	0.469	0.085***	1,782
	[0.498]	[0.499]	(0.025)	
Mother attended college	0.395	0.338	0.057**	1,783
	[0.489]	[0.473]	(0.025)	
Eligible for caste-based affirmative action	0.524	0.541	-0.016	1,789
	[0.500]	[0.499]	(0.025)	
Commitment to field of study	0.898	0.810	0.087***	1,597
	[0.304]	[0.392]	(0.019)	

Notes: This table reports summary statistics for key student characteristics and survey measures, disaggregated by gender. Column I reports mean values for female students, and Column II reports mean values for male students. Standard deviations are reported in brackets below the means in Columns I and II. Column III reports the gender difference in means, with robust standard errors in parentheses. Column IV reports the sample size for each variable. Mathematics and science proficiency, STEM-related anxiety, and STEM-related confidence are reported as z-scores (mean = 0, SD = 1). Commitment to field of study is measured by whether students report that they would choose the same branch or department if choosing again. Father attended college, mother attended college, eligibility for caste-based affirmative action, and commitment to field of study are binary measures. All rows report baseline measures collected at program entry. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 2: Balance checks for random assignment of students to classrooms.

Student characteristic	I Female faculty classroom	II Observations
Mathematics and science proficiency	0.009	13,714
	(0.030)	
STEM-related anxiety	-0.034	13,675
	(0.032)	
STEM-related confidence	0.010	13,675
	(0.033)	
Father attended college	-0.019*	28,787
	(0.011)	
Mother attended college	-0.011	28,807
	(0.009)	
Eligible for caste-based affirmative action	0.012	28,900
	(0.010)	
Female	0.011	28,949
	(0.010)	
P-value from F-test for joint significance	0.394	

Notes: This table reports balance tests for key student characteristics across classrooms taught by male and female faculty. Estimates are from linear regression estimated by OLS at the student–course level, with course fixed effects. Column I reports the difference in means of student characteristics between classrooms taught by female and male faculty. Column II reports the total number of observations for each characteristic. Standard errors clustered at the faculty level are in parentheses. Academic skills, STEM-related anxiety, and STEM-related confidence are reported as z-scores (mean = 0, SD = 1). The table also reports the P-value from an F-test for the joint significance of all characteristics. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 3: Design-based permutation diagnostics for random assignment to classrooms.

Panel A: Uniformity of per-classroom empirical P-values				
Baseline covariate	Strata	Classrooms	KS P-value	χ^2 P-value
Mathematics and science proficiency	209	712	0.236	0.090
STEM-related anxiety	209	712	0.414	0.353
STEM-related confidence	209	712	0.532	0.386
Panel B: Omnibus within-stratum permutation tests				
Baseline covariate	Strata	Classrooms	Size-weighted P-value	Standardized P-value
Father attended college	246	960	0.239	0.513
Mother attended college	246	960	0.855	0.177
Eligible for caste-based affirmative action	246	963	0.279	0.731
Female student	246	963	0.603	0.960

Notes: This table reports design-based permutation diagnostics using realized student-classroom rosters. Each diagnostic uses 1,000 placebo assignments that reassign students across classrooms within each course offering (department-course combinations corresponding to a unique institution and term) while preserving realized classroom sizes. Panel A reports tests of whether per-classroom empirical P-values are uniformly distributed for continuous baseline covariates. Panel B reports omnibus randomization P-values for binary covariates based on across-classroom dispersion within strata. The size-weighted statistic sums classroom-size-weighted dispersion around the stratum mean; the standardized statistic scales each stratum by its within-stratum variance. The number of strata and classrooms differs across panels because the permutation test is restricted to multi-classroom course offerings (excluding courses with a single classroom), and Panel A additionally restricts to students with non-missing baseline assessment and psychological measures.

Supplementary Table 4: Demographic characteristics of faculty by gender.

Faculty characteristic	I Female faculty	II Male faculty	III Difference	IV Observations
Professor	0.068	0.077	-0.009	430
			(0.026)	
Associate Professor	0.137	0.208	-0.071*	430
			(0.037)	
Experience in years	9.669	10.099	-0.430	429
			(0.669)	
Highest degree is PhD	0.279	0.366	-0.087*	431
			(0.047)	
Highest degree is master's or PhD in progress	0.687	0.616	0.071	431
			(0.048)	
Degree from elite college	0.336	0.324	0.012	430
			(0.048)	
Eligible for caste-based affirmative action	0.397	0.384	0.013	430
			(0.050)	

This table reports summary statistics for key demographic characteristics of faculty, disaggregated by gender. Column I reports mean values for female faculty, and Column II reports mean values for male faculty. Column III reports the gender difference in means, with robust standard errors in parentheses. Column IV reports the sample size for each variable. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 5: Effect of female faculty on female and male students' academic achievement.

	I Course grade	II Course grade
Female faculty × Female student (β_1)	0.070**	0.070**
	(0.030)	(0.030)
Female faculty (β_2)	0.004	0.007
	(0.027)	(0.027)
$\beta_1 + \beta_2$	0.074**	0.077**
	(0.032)	(0.031)
P-value ($H_0 : \beta_1 + \beta_2 = 0$)	0.020	0.012
Student fixed effects	Yes	Yes
Course fixed effects	Yes	Yes
Classroom fixed effects	No	No
Faculty-level controls	No	Yes
Student-level controls	No	No
Observations	28,878	28,824

This table reports auxiliary estimates for course grades using Equation (2). The outcome is the course-grade z-score (mean = 0, SD = 1), and the unit of observation is a student–course pair. Estimates are from linear regression estimated by OLS with student fixed effects and course fixed effects. The specification omits classroom fixed effects so that the main effect of female faculty can be estimated; the preferred course-grade specification remains Table 1. Standard errors (in parentheses) are clustered at the faculty level. β_1 estimates the additional effect of female faculty on female students relative to male students. β_2 captures the effect of female faculty on male students. $\beta_1 + \beta_2$ is the total effect of female faculty on female students. Column II adds faculty-level controls for age, academic rank, years in higher education, PhD status, reservation status, and whether the faculty member attended an elite college. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 6: Robustness of the effect of female faculty on female students' academic achievement.

	I Course grades
Sample restricted to non-elite colleges (N = 26,786)	
Female faculty × Female student	0.083**
	(0.032)
Sample restricted to STEM courses (N = 26,258)	
Female faculty × Female student	0.087***
	(0.033)
Equally weighted classrooms (N = 28,376)	
Female faculty × Female student	0.090**
	(0.040)
Using course ranks as outcome (N = 28,558)	
Female faculty × Female student	2.638***
	(0.934)
Student fixed effects	Yes
Classroom fixed effects	Yes
Student-level controls	No
Faculty-level controls	No

This table reports robustness checks for the effect of female faculty on female students' academic outcomes. All coefficients are estimated from Equation (1) with course-grade outcomes. Estimates are from linear regression estimated by OLS. The unit of observation is a student–course pair, and standard errors (in parentheses) are clustered at the faculty level. The first panel restricts the sample to non-elite colleges. The second panel restricts the sample to STEM courses. The third panel weights classrooms equally to account for differences in class size. The fourth panel uses course ranks as outcomes instead of z-scored grades. All specifications include student and classroom fixed effects. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 7: Effect of female faculty exposure on female students' academic achievement by commitment to field of study.

	I Effect on female student (SD)	II Observations
Panel A: Course grades		
Would change field	0.305***	4,142
	(0.088)	
Would choose same field	0.049	21,756
	(0.033)	
Student fixed effects	Yes	
Classroom fixed effects	Yes	
Panel B: Standardized test scores		
Would change field	0.025	266
	(0.041)	
Would choose same field	0.032*	1,335
	(0.017)	
Department fixed effects	Yes	
Subject fixed effects	Yes	
Student-level controls	Yes	
Faculty-level controls	Yes	

This table reports estimates of the effect of female faculty exposure on female students' academic outcomes, separately by whether students reported at program entry that they would change their branch or department if choosing again. We refer to this measure as commitment to one's field of study. Panel A reports coefficients from linear regression estimated by OLS using Equation (1) with course-grade z-scores (mean = 0, SD = 1) as the outcome; the specification includes student fixed effects and classroom fixed effects, the coefficient measures the effect of being taught by a female instructor, and standard errors (in parentheses) are clustered at the faculty level. Panel B reports coefficients from linear regressions estimated by OLS using Equation (3) with standardized mathematics and science test scores (standardized within subject) as the outcome; the specification includes department/school fixed effects, subject fixed effects, student-level controls, average faculty-level controls, and baseline test score controls. The coefficient measures the effect of a 10-percentage-point increase in the share of courses taught by female faculty, and standard errors are clustered at the student level. Standardized tests were administered to a randomly selected half of the sample; tested students appear twice (once in mathematics and once in science). Approximately 10% of students did not respond to the commitment question at program entry. Column I reports the coefficient estimate, and Column II reports the number of observations. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 8: Effect of female faculty on female students' course grades by prior achievement, STEM-related confidence, and STEM-related anxiety.

	I Female faculty × Female student	II Observations
Prior academic achievement		
Low	0.120**	5,766
	(0.055)	
High	0.037	7,552
	(0.051)	
STEM-related confidence		
Low	0.113**	6,646
	(0.056)	
High	0.041	6,653
	(0.053)	
STEM-related anxiety		
Low	0.057	6,630
	(0.055)	
High	0.081	6,648
	(0.050)	
Student fixed effects	Yes	
Classroom fixed effects	Yes	
Student-level controls	No	
Faculty-level controls	No	

This table reports estimates of the effect of female faculty on female students' course-grade z-scores (mean = 0, SD = 1) by subgroups. Smaller sample sizes reflect that these measures were collected alongside the standardized mathematics and science assessments, which were administered to a randomly selected half of the sample. Column I reports the interaction effect of female faculty and female student from linear regression estimated by OLS using Equation (1); the unit of observation is a student–course pair, and standard errors (in parentheses) are clustered at the faculty level. Column II reports the number of observations per subgroup. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 9: Effect of female faculty on commitment to field of study and classroom engagement.

	I Commitment to field of study	II Engagement in class
Female faculty share × Female student	0.008	0.025*
	(0.010)	(0.013)
Female faculty share	0.007	-0.036**
	(0.013)	(0.017)
Female student	-0.015	-0.101**
	(0.041)	(0.051)
Mean	0.859	0.698
Department fixed effects	Yes	Yes
Student-level controls	Yes	Yes
Faculty-level controls	Yes	Yes
Observations	1,748	1,747

This table reports estimates of the effect of female faculty exposure on female students' commitment to their field of study (Column I) and engagement in class (Column II). Female faculty share represents the proportion of a student's courses taught by female faculty over two years and is scaled by a factor of 10. Commitment to field of study is measured by whether students report that they would choose the same branch or department if choosing again. Classroom engagement is defined as asking questions at least once every few classes. Estimates are from linear probability model estimated by OLS at the student level. All specifications include department/school fixed effects, student-level controls, average faculty-level controls, and the baseline value of the corresponding outcome. Robust standard errors are in parentheses. Sample means for each outcome are reported at the bottom. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 10: Effect of male peers' stereotypical beliefs on female students' outcomes (2SLS).

	I STEM-related confidence	II STEM-related anxiety
Male peers' stereotypical beliefs (pp)	-0.034	0.049*
	(0.028)	(0.027)
First-stage coefficient (on instrument)	0.084***	0.086***
	(0.022)	(0.021)
Reduced-form coefficient (on instrument)	-0.003	0.004**
	(0.002)	(0.002)
Kleibergen–Paap rk Wald F-statistic	14.74	16.11
Anderson–Rubin P-value	0.200	0.038
Observations	526	526
Student fixed effects	Yes	Yes
Subject fixed effects	Yes	Yes

This table reports two-stage least squares estimates of the effect of exposure to male peers' stereotypical beliefs on STEM-related confidence (Column I) and STEM-related anxiety (Column II) among female students. The endogenous regressor is the percentage-point share of male peers endorsing the gender stereotype. The instrument is male classmates' leave-one-classroom-out exposure to female instructors in their other first- and second-year courses, interacted with a mathematics-versus-science subject contrast $\{+1, -1\}$. All specifications include student fixed effects, subject fixed effects, and the matching baseline outcome; outcomes are standardized within subject (mean = 0, SD = 1). Robust standard errors clustered at the student level are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 11: Effect of female faculty on interactions with students in class and office hours.

	I Call on student in class	II Discussion during breaks	III Office hours
All students			
Female faculty share × Female student	-0.002	-0.017**	-0.011
	(0.009)	(0.008)	(0.009)
Female faculty share	-0.003	0.014**	0.005
	(0.006)	(0.006)	(0.006)
Female student	-0.006	0.082**	0.005
	(0.038)	(0.034)	(0.039)
Mean	0.705	0.766	0.637
Observations	1,749	1,742	1,746
Students with low prior performance			
Female faculty share × Female student	0.002	-0.008	0.020
	(0.014)	(0.010)	(0.014)
Female faculty share	0.003	0.013	-0.008
	(0.009)	(0.008)	(0.010)
Female student	-0.010	0.092*	-0.010
	(0.059)	(0.049)	(0.058)
Mean	0.733	0.799	0.664
Department fixed effects	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes
Student-level controls	Yes	Yes	Yes
Faculty-level controls	Yes	Yes	Yes
Observations	692	688	689

This table reports estimates of the effect of female faculty on interactions with students in different academic settings. Outcomes take a value of 1 if the student engaged in the respective interaction and 0 otherwise. Estimates are from linear probability model estimated by OLS. The unit of observation is a student–subject pair (mathematics or science). All specifications include department/school fixed effects, subject fixed effects, student-level controls, average faculty-level controls, and the baseline value of the corresponding interaction. Standard errors (in parentheses) are clustered at the student level. Female faculty share represents the proportion of a student's courses taught by female faculty over two years and is scaled by a factor of 10. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 12: Effect of faculty gender on teaching-related activities and teaching practices.

	I Female faculty	II Mean	III Observations
Hours spent weekly on:			
Tutoring	0.077	2.761	491
	(0.205)		
Homework and tests	0.005	3.724	491
	(0.231)		
Course-related work	0.029	2.771	491
	(0.263)		
Advising students	-0.671**	3.314	490
	(0.283)		
Lesson planning	-1.087**	7.176	491
	(0.498)		
Teaching classes	-1.482**	11.329	491
	(0.678)		
Teaching Practices Inventory scores			
In-class activities	-0.164	0	495
	(0.103)		
Assignments	0.229**	0	495
	(0.105)		
Feedback and testing	-0.004	0	495
	(0.094)		
Collaboration	-0.047	0	495
	(0.091)		
Department fixed effects		Yes	
Faculty-level controls		Yes	

This table reports differences between male and female faculty in time spent on teaching-related activities and use of teaching practices. Estimates are from linear regression estimated by OLS at the faculty-department level, with department/school fixed effects and faculty-level controls for academic rank, years in higher education, PhD status, reservation status, and whether the faculty member attended an elite college. Column I reports the estimated difference for female faculty relative to male faculty. Column II reports the mean of the dependent variable, and Column III reports the number of observations. The first panel reports differences in weekly hours spent on teaching-related activities. The second panel reports differences in Teaching Practices Inventory (TPI) scores, which measure different aspects of instructional practices. TPI scores are reported as z-scores (mean = 0, SD = 1). Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 13: Effects of female faculty with stereotypical and non-stereotypical beliefs on female students' academic achievement.

	I Course grade (mathematics and science)	II Course grade (other courses)
Beliefs about female student ability (at college entry)		
Female faculty with non-stereotypical beliefs × Female student	0.083**	
	(0.034)	
Female faculty with stereotypical beliefs × Female student	0.081	
	(0.057)	
Share of female faculty with stereotypical beliefs	0.245	
Share of male faculty with stereotypical beliefs	0.398	
Beliefs about female student ability (after two years of college)		
Female faculty with non-stereotypical beliefs × Female student	0.087***	0.070**
	(0.033)	(0.034)
Female faculty with stereotypical beliefs × Female student	0.037	0.082*
	(0.052)	(0.044)
Share of female faculty with stereotypical beliefs	0.177	0.245
Share of male faculty with stereotypical beliefs	0.275	0.306
Student fixed effects	Yes	Yes
Classroom fixed effects	Yes	Yes
Student-level controls	No	No
Faculty-level controls	No	No
Observations	28,053	27,914

This table reports estimates of the impact of female faculty with stereotypical or non-stereotypical beliefs on female students' course-grade z-scores (mean = 0, SD = 1). Estimates are from linear regression estimated by OLS at the student–course level, with student fixed effects and classroom fixed effects. Faculty gender beliefs indicate whether faculty perceive male and female students as at least equally capable (non-stereotypical) or believe male students are more qualified (stereotypical). These beliefs were elicited for mathematics and science (Column I) and for other major-specific courses (Column II) separately. Standard errors (in parentheses) are clustered at the faculty level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 14: Effects of female faculty with fixed and growth mindset on female students' academic achievement.

	I Course grade
Mindset about malleability of intelligence	
Female faculty with growth mindset × Female student	0.108**
	(0.044)
Female faculty with fixed mindset × Female student	0.067*
	(0.037)
Share of female faculty with fixed mindset	0.562
Share of male faculty with fixed mindset	0.644
Mindset that skills can be learned, but intelligence is fixed	
Female faculty with growth mindset × Female student	0.071*
	(0.043)
Female faculty with fixed mindset × Female student	0.085**
	(0.037)
Share of female faculty with fixed mindset	0.575
Share of male faculty with fixed mindset	0.559
Student fixed effects	Yes
Classroom fixed effects	Yes
Student-level controls	No
Faculty-level controls	No
Observations	28,211

This table reports estimates of the impact of female faculty with fixed or growth mindset on female students' course-grade z-scores (mean = 0, SD = 1). Estimates are from linear regression estimated by OLS at the student–course level, with student fixed effects and classroom fixed effects. The excluded category is Male faculty × Female student. In the first panel, fixed mindset is defined as endorsing the belief that intelligence is largely innate and unchangeable; all other responses are categorized as growth mindset. In the second panel, fixed mindset reflects the belief that intelligence is largely innate and unchangeable but new skills can be learned; growth mindset includes all other responses. Standard errors (in parentheses) are clustered at the faculty level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 15: Effects of female faculty by academic rank and doctoral degree on female students' academic achievement.

	I Course grade
Academic rank	
Female faculty, not professor or asst. professor × Female student	0.098***
	(0.037)
Female professor or asst. professor × Female student	0.050
	(0.042)
Share of female faculty with rank professor or assistant professor	0.279
Share of male faculty with rank professor or assistant professor	0.366
Doctoral degree	
Female faculty without PhD × Female student	0.100***
	(0.037)
Female faculty with PhD × Female student	0.047
	(0.043)
Share of female faculty with PhD	0.211
Share of male faculty with PhD	0.285
Student fixed effects	Yes
Classroom fixed effects	Yes
Student-level controls	No
Faculty-level controls	No
Observations	28,558

This table reports estimates of the impact of female faculty, by rank and doctoral degree status, on female students' course-grade z-scores (mean = 0, SD = 1). Estimates are from linear regression estimated by OLS at the student–course level, with student fixed effects and classroom fixed effects. The excluded category is Male faculty × Female student. Standard errors (in parentheses) are clustered at the faculty level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 16: Average direct and indirect effects from descriptive mediation analysis (standardized test scores).

	I STEM-related confidence	II STEM-related anxiety
Average indirect treatment effect	0.000	0.003
	(0.001)	(0.002)
Average direct treatment effect	0.017	0.019
	(0.016)	(0.017)
Average treatment effect	0.017	0.022
	(0.017)	(0.017)
Proportion mediated (%)	–	14.4
Student-level controls	Yes	Yes
Faculty-level controls	Yes	Yes
Baseline outcome control	Yes	Yes
Baseline mediator control	Yes	Yes
Subject indicator	Yes	Yes
Observations	520	520

This table reports average direct and indirect effects from a descriptive mediation analysis with standardized test scores as the outcome. Effects are reported in standard deviation units of standardized test score and are evaluated for a 10-percentage-point increase in female faculty share. Estimates are restricted to female students. Estimates are from descriptive mediation models. The unit of observation is a student–subject pair among female students, and effects are reported for a 10-percentage-point increase in female faculty share. Models include controls for average faculty characteristics and student characteristics, baseline outcome and baseline mediator, and an indicator for mathematics or science test. Robust standard errors clustered by student are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 17: Survey items used to construct STEM-related anxiety and confidence

Item	Statement
1	I often worry about the challenges involved in learning [subject].
2	I am nervous when doing homework for [subject] courses.
3	I get good grades in [subject].
4	I am nervous when doing problem sets for [subject] courses.
5	I am a fast learner when it comes to [subject].
6	I think [subject] is one of my best subjects.
7	I often feel lost when working on problem sets for [subject] courses.
8	When taking classes in [subject], I can understand even the most difficult parts of the content.
9	I worry that my grades in [subject] courses will be bad.

Notes: This table reports the survey items used to construct the STEM-related anxiety and STEM-related confidence indices. Items were administered separately for mathematics and science (physics) at program entry and approximately two years later. The survey prompt read: "We are interested in your subjective experience in [subject]. Please indicate whether or not you agree with the following statements," where [subject] was replaced by "mathematics" or "physics." The response scale for all items was: 1 = strongly agree, 2 = agree, 3 = disagree, and 4 = strongly disagree. The STEM-related anxiety index uses Items 1, 2, 4, 7, and 9, which capture students' worries about learning and performance and their feelings of being nervous or lost when completing coursework. The STEM-related confidence index uses Items 3, 5, 6, and 8, which capture students' self-perceived ability, including whether they get good grades, view themselves as fast learners, see the subject as a strength, and feel capable of understanding difficult content. The anxiety items were adapted from the OECD PISA Mathematics Anxiety scale; the confidence items were adapted from OECD PISA Mathematics Self-Concept items, with parallel wording for physics [7, 8]. Items were recoded as needed before index construction so that higher values correspond to higher STEM-related anxiety or higher STEM-related confidence, respectively.

Supplementary Table 18: Effect of female faculty on female students' STEM-related confidence and anxiety using principal component measures.

	I STEM-related confidence	II STEM-related anxiety
Female faculty share × Female student	-0.011	-0.050***
	(0.015)	(0.016)
Female faculty share	-0.005	0.004
	(0.013)	(0.014)
Female student	0.014	0.044
	(0.064)	(0.069)
Department fixed effects	Yes	Yes
Subject fixed effects	Yes	Yes
Student-level controls	Yes	Yes
Faculty-level controls	Yes	Yes
Observations	1,568	1,562

This table reports the effects of female faculty on students' STEM-related confidence and anxiety using principal component measures. Column I reports results for confidence, and Column II reports results for anxiety, both measured as z-scores (mean = 0, SD = 1) of the first principal component. Estimates are from linear regression estimated by OLS using Equation (3). The sample consists of the randomly selected half of students who completed standardized assessments in mathematics and science. The unit of observation is a student–subject pair (mathematics or science). The specification includes the baseline value of the corresponding outcome, student-level controls, average faculty-level controls, department/school fixed effects, and subject fixed effects. Standard errors (in parentheses) are clustered at the student level. Female faculty share is scaled by a factor of 10. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Table 19: Timing and data sources of variables used in the analysis

Variable	Data source and timing
Course grades	Administrative records
Female faculty exposure	Administrative records
Standardized test scores	Baseline and endline student assessments
STEM-related anxiety and confidence	Baseline and endline student survey
Commitment to field of study	Baseline and endline student survey
Classroom engagement and faculty-student interactions	Baseline and endline student survey
Student gender beliefs	Endline student survey
Student demographics	Baseline student survey
Faculty demographics	Endline faculty survey
Faculty gender beliefs	Endline faculty survey
Faculty mindset	Endline faculty survey
Teaching-related activities and practices	Endline faculty and student surveys

Notes: This table summarizes the timing and data sources of variables used in the analyses. Baseline measures were collected at the start of the program, and endline measures were collected after two years. Administrative records cover course enrollments, classroom assignments, instructor assignments, and grades during first- and second-year coursework.

Supplementary References

References

[1] Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses*. Springer, New York, 3 edition, 2005.

[2] Phillip I. Good. *Resampling Methods: A Practical Guide to Data Analysis*. Birkhäuser, Boston, 3 edition, 2006.

[3] Scott E. Carrell and James E. West. Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3):409–432, 2010. doi: 10.1086/653808.

[4] Jaegeum Lim and Jonathan Meer. Persistent effects of teacher–student gender matches. *Journal of Human Resources*, 55(3):809–835, 2020. doi: 10.3368/jhr.55.3.0218-9314R4.

[5] Kosuke Imai, Luke Keele, and Dustin Tingley. A general approach to causal mediation analysis. *Psychological Methods*, 15(4):309–334, 2010. doi: 10.1037/a0020761.

[6] Tyler J. VanderWeele. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, New York, NY, 2015. ISBN 978-0-19-932587-0.

[7] Organisation for Economic Co-operation and Development. PISA 2012 student questionnaire (English–USA). `https://www.oecd.org/content/dam/oecd/en/data/datasets/pisa/pisa-2012-datasets/main-survey/questionnaires/Student%20Questionnaire%20Form%20B%20PISA%202012.pdf`, 2012.

[8] Organisation for Economic Co-operation and Development. *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing, Paris, 2023. doi: 10.1787/53f23881-en.