Southern Economic Journal

WILEY | SEA SOUTHERN ECONOMIC ASSOCIATION

RESEARCH ARTICLE OPEN ACCESS

Fixing the Leaky Pipeline: Affirmative Action in Local Elite Colleges and Subject Choice

Ritika Gupta

Department of Economics, Wageningen University, Wageningen, the Netherlands

Correspondence: Ritika Gupta (ritika.gupta@wur.nl)

Received: 19 February 2025 | **Accepted:** 22 October 2025

Keywords: affirmative action | elite college | gender | STEM | subject choice

ABSTRACT

Women are largely underrepresented in STEM careers associated with higher labor market returns. This gender gap is even more stark in a context where societal biases are prevalent and female role models are lacking. This paper investigates the impact of an affirmative action (AA) policy implemented in an elite educational institution in India that ensures additional seats specifically for women in undergraduate STEM courses. After the policy was implemented, the proportion of women enrolling increased by 100%, proportion of women taking the college entrance exam increased by 10% and those qualifying the exam increased by 15%. Using nationally representative data, I employ a triple difference strategy and find a 27% increase in the probability of studying science courses after Grade 10 amongst younger girls exposed to this policy, suggesting increase in the future expected earnings of women.

JEL Classification: I24, I25, I28, O15, J16

1 | Introduction

The under-representation of women in Science, Technology, Engineering and Mathematics (STEM) fields—an outcome of the progressive loss of women in STEM or the "leaky pipeline"—is recognized as one of the major causes of the gender wage gap and occupational segregation (Sharpe 1976; Wolpe 1978; Daymont and Andrisani 1984; Beede et al. 2011; Deem 2012; Resmini 2016). Deep-rooted gender norms and the lack of role models hinder the narrowing of the gender gap in STEM—which, if achieved, can lead to an increase of \$12–\$28 trillion in the global economy via increased labor market activity and productivity of women, according to a recent research report by McKinsey (Maceira 2017; Munoz-Boudet and Revenga 2017; Woetzel et al. 2020).

One set of policies that aim to narrow this gap involves affirmative action (AA) often developed and employed by educational institutions to break entry barriers (Ceci and Williams 2015; Bastarrica et al. 2018). Whether programs like these can influence the career path of women in male-dominated fields is a first-order empirical question. On one hand, these are meant to encourage women by increasing their likelihood of entry; but on the other, they can reinforce stereotypes and gender roles (Matheson et al. 1994).

In this paper, I analyze one such program introduced at an elite tier of engineering colleges in India—Indian Institute of Technology (IIT)—that reserved extra seats for women at every new undergraduate STEM course cohort entering an IIT campus in 2018. I investigate the impact of the policy on subject choice pursued by girls after completing Grade 10, by exploiting the exogenous variation in proximity to these institutions in a context where students prefer going to college closer to their homes. The presence of at least one IIT campus in almost every state in India provides large spatial variation in the proximity to the institute. Admission to an IIT is based purely on merit, eliminating any migration or selective sorting patterns that could arise from the knowledge of this policy.

The educational system in India requires students to choose one of three tracks—Science, Commerce, or Humanities—after completing Grade 10 in school, which then defines the courses of study in subsequent grades (Grades 11 and 12). In order to choose a STEM major at an Indian university, a prospective student is required to have studied subjects under the Science track in Grades 11 and 12. In particular, getting an admit into an IIT requires qualifying for a very selective entrance examination that tests knowledge of science track courses—Physics, Chemistry, and Mathematics. As a result, subject choice in high school defines one's career path to a large extent. Any policy, therefore, that can influence choice at this stage can increase the likelihood of advancing into a STEM career. I analyze this subject choice in a context specific but not limited to India.[1]

After the implementation of this policy, the proportion of girls at IITs nearly doubled from 8.7% to 16%, resulting in an average yearly increment of over 1300 seats across all IIT campuses. The proportion of women taking the IIT entrance exam increased by 10.5%, and those qualifying for the exam increased by about 15.3%. Overall, this translates to about 4 more females taking the exam and 1 more female clearing the exam for each extra seat added to the seat pool. Since the policy added new seats and reserved them for women, they didn't displace boys. In fact, there is a marginal increase in the absolute number of males enrolled at IITs, although it is insignificant.

Students living closer have a comparative advantage in responding to the policy over those living farther as long travel times increase safety concerns. This coupled with strict social norms strongly influences education decisions, especially for girls. I exploit the fact that the preference for an educational institute closer to one's home is salient in India. Moreover, conditional on clearing the IIT entrance exam, students indicate their preferred IIT campus and engineering field. However, the location of some IIT campuses in remote areas (Kharagpur, Roorkee, Guwahati etc.) can constrain women's choice set (Borker 2017). Stereotypes associated with certain fields (such as Mechanical or Civil engineering) being "masculine" (Chanana 2007) further limits women's choices, making it difficult for them to enroll in an elite college and instead making them settle for a lower quality college in closer proximity. In light of the aforementioned context, I evaluate the policy for the "marginal" girl living close to an IIT campus who faces weaker safety and transport barriers.

I build a conceptual framework to illustrate the trade-off faced by a girl when deciding subjects to study in Grade 11. The benefit of studying science is twofold - (1) higher wage premium associated with the science track and (2) possibility of studying STEM at an elite college (EC) and earning the EC wage premium. The cost of studying science is represented by the distaste for the subject, reflecting the gendered stereotype. There is also an additional cost of travel. The framework predicts that for a girl who lives far from an EC, such that the wage premium is not high enough so as to outweigh the extra cost of travel, she will not go to EC and pursue her subject of study from the local college (LC) depending on the distaste for science. For a girl who lives close enough to EC, as long as her distaste for science is low, she will choose to study STEM there if selected. An AA policy such as supernumerary seats can potentially increase the likelihood of entering EC and thus influence girls closer to ECs and make them switch from a non-science to science subject.

To empirically estimate the impact of the policy, I use nationally representative cross-section data collected in 2017–2018 from a special round of the National Sample Survey (NSS) focusing on education and estimate the effect of the announcement of the policy. I compare subjects pursued after Grade 10 for cohorts making their decisions before and after the policy was announced (first difference). I calculate a triple difference estimate which compares the first difference between girls and boys living close to IIT campuses with those that live far. The key identifying assumption in my model is that conditional on district-specific characteristics and individual-level controls, if the policy would not have been introduced, gender gaps in science between close and far districts from an IIT would follow parallel trends. I test this assumption by testing for differential trends in the older cohorts. I fail to reject the parallel trends assumption for the triple difference estimate.

In order to create the spatial variation, I first find distances of each IIT from all districts. Districts that lie within a 30 km radius of an IIT are considered "close", whereas districts that lie outside that radius but within 200 km are considered "far". I use 30 km as the threshold, as it is a reasonable distance that can be commuted on a daily basis.[2] Moreover, IITs exempt students from living on campus as long as they reside within 30 km of an IIT. The fact that I study choices that students make when they are in school, that is at a time when they reside at home with their parents, alleviates any selective sorting or migration issues, as it is unlikely that individuals will change their residence with this policy announcement. One particular concern in considering districts farther from IITs as controls is that these areas can be quite different from the areas closer to these colleges. I address this concern by using synthetic difference-in-difference (SDID) weights for each district and age category. This is done by running a synthetic DID (Arkhangelsky et al. 2021) specification on a collapsed district-age panel. Considering the "close" districts as treated, I find unit-specific weights for "far" districts and time-specific weights for three age brackets. This allows me to re-weight my original regression to match trends in science in treated and control districts.

The key result of the paper is that the inclusion of supernumerary seats for girls is associated with a 6.7 percentage point increase (about 27% of the baseline average population of women studying science) in the likelihood of choosing the science track after Grade 10 in areas closer to IIT campuses. I find a similar effect when the regression is weighted using SDID. The estimate suggests that the likelihood of choosing the science track increased by around 0.02% for every additional seat that was added. If we were to assume that studying the science track increases expected earnings by 22% for females in India (as it does for males according to Jain et al. 2022), then AA at elite engineering colleges has the potential to increase the expected earnings of women by about 6% and therefore can have huge implications in narrowing the gender wage gap.

I also analyze the impact of the policy on other education outcomes. I do not find any effect on educational attainment, private coaching uptake, or other expenditure in education. I perform heterogeneity analysis to see whether the increase in science is driven by sibling spillover effects or parents' education level, but I do not find such evidence. Lastly, I perform a variety of sensitivity checks on the key result, and the main results are robust.

The paper adds to the broad AA literature and attempts to exploit policy variation to study educational outcomes of women, and STEM in particular. The majority of the studies on college-based AA have focused on upstream effects by looking at post college-entry outcomes (enrolment, completion, and performance) of AA beneficiaries. Caste-based reservations have been studied in India to determine their targeting and matching properties and to assess college performance of lower-caste students (Bertrand et al. 2010; Bagde et al. 2016; Aygün and Turhan 2017). Other studies have looked at racial differences in college enrolment and attainment of minorities by focusing on AA bans (Backes 2012; Hinrichs 2012; Arcidiacono et al. 2016; Bleemer 2022). My paper distinguishes itself by focusing on the downstream effect of college-level AA for women and, in particular, studies their subject choice in school. A similar study also looks at the downstream effects of reservations in government jobs for lower-caste individuals but focuses mostly on school and grade completion as outcomes (Khanna 2020). Moreover, studies on AA for women are limited to job domains such as politics (Beaman et al. 2009); corporate board leadership (Matsa and Miller 2013), and law enforcement (Miller and Segal 2019; Sukhtankar et al. 2022). This paper distinguishes itself from other related papers by studying the impact of college-based AA on pre-college outcomes of women.

This paper contributes to the literature on subject choice, which has implications for labor market earnings and the gender wage gap. Previous literature has established the role of subject choice as an important contributor to the gender wage gap. Men are more likely to be STEM-ready before college by scoring higher on science tests and having taken more advanced math and science courses, which accounts for 35% of the overall gender gap in STEM careers (Speer 2023). This gender gap in major choice strongly predicts gender wage gaps—even when accounting for occupation choice (Sloane et al. 2019). In the US and Canada, the gender gap in the likelihood of graduating with a STEM-related degree explains about 20% of the wage gap between younger college-educated men and women (Card and Payne 2021). Jain et al. (2022) establish that conditional on ability, choosing the science track in high school generates 22% greater earnings for Indian males. Even though their study cannot provide enough evidence for females due to low female labor force participation in India, the science track is also likely to be associated with higher wages for females. Moreover, earnings associated with elite public colleges are much higher than other colleges (Zimmerman 2019; Sekhri 2020). This paper investigates whether this choice at high school can be influenced for girls by an elite public college policy that can potentially reduce the gender wage gap. The paper also adds to a wide literature that looks at a variety of factors which determine the choice of subject such as ability, earnings, tastes, and preferences (Wiswall and Zafar 2015), peer effects (Fischer 2017), siblings spillovers (Altmejd et al. 2021), and role models (Porter and Serra 2020).

This paper also contributes to the literature on the gender gap in STEM enrolment and the "leaky pipeline" by analyzing a policy that has the potential to narrow that gap. Previous studies have identified the existence of the gender gap in math and participation in STEM fields (Fryer Jr and Levitt 2010; Adams and Kirchmaier 2016) as well as the gender gap in higher secondary subject choice in India (Sahoo and Klasen 2021). Other studies have tried to explain this gap by analyzing gender differences in test-taking behavior in a competitive environment (Niederle and Vesterlund 2010; Buser et al. 2014, 2017) and establishing the role of culture in determining math performance (Nollenberger et al. 2016). While this paper does not establish or explain gender differences in STEM, it attempts to evaluate a policy to answer if it can fix the "leaky pipeline" by expanding college opportunities for women.

2 | Context and Background

Indian Institutes of Technology (IIT) are public engineering and research institutions in India and are ranked highest in India. As of 2020, there were 23 IITs located across the country, each of which is autonomous but administered through a common IIT council.[3] The most common, competitive, and sought-after degree at IITs is Bachelor of Technology (B.Tech).[4] To seek admission to one of these B.Tech programs, students are required to pass a competitive entrance examination covering topics from subjects taught in the Science track in Grades 11 and 12. It focuses on the application of concepts through novel questions in a stipulated time frame, which makes it one of the hardest exams to crack. The highest scorers are admitted into one of the IITs based on their rank and their declared field and location preference. Every year, 1.5 million students take the exam and apply to the undergraduate programs for which only about 16,000 seats are available across all IITs. Conditional on having studied Science in Grades 11–12 at school, IITs, therefore, admit students purely on the basis of their performance in the entrance exam (and therefore on merit).

2.1 | Supernumerary Seats for Women

In terms of the structure and eligibility of admissions to IITs, there are no barriers whatsoever against women in applying. Yet there are large gender differences in application, admission, and entry of women generating an acute under-representation of women in undergraduate engineering courses in IITs. Gupta (2020) mentions that in 2016 only 19% of the candidates writing the entrance exam were women. Out of the candidates who passed, only 12.5% were women, and finally, there were only 8% women in the incoming cohort of students across all the IITs. The gender ratio at IITs has been highly skewed since their inception, and this very low proportion of women has led to the introduction of supernumerary seats in 2018. The agenda of the policy is to create new seats for women in every undergraduate program at every IIT until a minimum percentage of female enrolment is achieved.

2.2 | Trends in Enrolment

I first study the trends in enrolment of women in the undergraduate programs in IITs before and after the policy came into effect in 2018. For this purpose, I utilize the Annual Reports available for 20 IITs on their website to gather yearly data on total new admissions in the 4-year B.Tech degree programs.

Prior to 2018, each campus on average admitted 350 boys and 33 girls in the new cohort every year. After 2017, these institutes have been admitting 412 boys and 90 girls on average.[5] Figure 1 presents the distribution of newly admitted students by gender averaged over IITs using the data for academic years between 2014–2015 and 2021–2022. There is a clear change in trend after 2017.

I use the IIT-year panel to plot these trends and measure the impact of the policy on the gender ratio and the proportion of women at IITs. I report robust standard errors clustered at each IIT. In particular, I run the following fixed-effects regression:

$$y_{it} = \alpha + \beta Post_t + \mu_i + \epsilon_{it} \tag{1}$$

where y_{it} is the gender-ratio or the proportion of women in IIT i in year t; μ_i are IIT fixed effects that capture any time-invariant IIT-specific characteristics and $Post_t$ is a dummy which takes the value 1 for the years that included supernumerary seats for women (i.e., after 2017). As presented in Table 1, we see that on average, the introduction of the policy led to a 11 percentage point increase in the gender ratio and a 8.7 percentage point increase in the proportion of women at IITs. These estimates are statistically significant and suggest that compared to the average baseline, the proportion of girls enrolled in these IITs have nearly doubled. Based on the average enrolment numbers, the absolute number of girls in a cohort increased from 759 to 2070, which is an increase of about 1300 girls. Year-specific coefficients plotted in Figure 2 depict the trend of female enrolment at IITs.

2.3 | Trends in Applications

The increase of women at IITs is an outcome of the policy implementation. In order to investigate if more women are also applying to IITs (or taking the IIT entrance exam), I collect data on applicants from the IIT entrance exam annual reports. The IIT campuses are divided into seven regional zones. Every year, one of these seven IIT zones conducts the exam and publishes the exam statistics in a report available on their website. I utilize these reports to gather the number of applicants and qualified students. A total of 33,307 girls and 138,506 boys registered (or applied) for the IIT entrance exam in 2017. The corresponding number for those who qualified for a seat at IIT was 7259 girls and 43,781 boys.

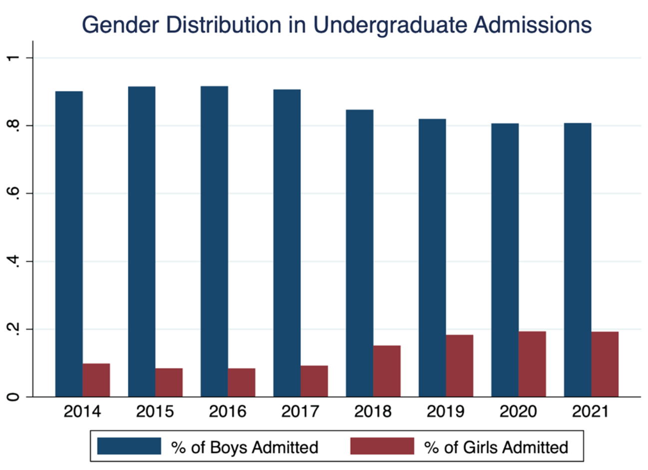


FIGURE 1 | Gender distribution in undergraduate admissions at IITs. [Color figure can be viewed at wileyonlinelibrary.com]

I create an IIT zone-year panel and run the fixed effects regression as in the previous section. As shown in Table 2, I observe that there has been an increase of nearly 2 percentage points in the proportion of women who take the IIT entrance exam as well as in the proportion of women clearing the exam, suggesting a 10.5% increase in the proportion of women taking the exam and a 15.3% increase in the proportion of women qualifying the exam. In absolute terms, this translates to a yearly average of 4179 more female registrations and 1080 more females qualifying the exam. For every additional seat added in IITs for women, 4 more females take the IIT entrance exam and 1 more female qualified the exam.[6]

3 | Conceptual Framework

Consider a simple framework where a girl after finishing Grade 10 decides whether to choose a science track or a non-science track. I denote the labor market return from studying science as α_s which I assume to be strictly greater than the labor market return from non-science, α_n, as science track is associated with higher labor market earnings (Jain et al. 2022). However, there is distaste associated with studying science which denotes the notion that science is "bad" for girls. I assume idiosyncratic distaste for studying science, δ_i.

After passing Class 12 in the track she studied in, she proceeds to study in the university. There are two universities—Local (L) and Elite (E). L offers all courses and by definition, is close to the girl's home. E offers only STEM courses and there is wage premium, $\gamma_E > 0$, associated with studying in E. I assume that the girl does not drop out of education before going to college. The choice of studying at University E only becomes available if the girl studies the science track at school and passes the competitive entrance exam to get admission into E. The probability of passing the exam is p_E which I assume is same for everyone. University L, on the other hand, is always open to admission and she can always join it irrespective of whether she gets admission in E or not. I, therefore, assume for simplicity that the probability of attending L is 1.[7] In

TABLE 1 | Impact on female enrolment at IITs.

	Gender ratio	Proportion of women
Post	0.116***	0.0877***
	(0.00823)	(0.00670)
Constant	0.100***	0.0893***
	(0.00379)	(0.00309)
Observations	128	128
R^2	0.720	0.725
Number of IITs	20	20

Note: Data Source is Annual Reports of 20 IITs. Gender ratio is the number of females divided by the number of males. Proportion of women is defined as the number of females divided by the total number of students admitted. Independent variable is a dummy taking value 1 for post-policy years.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

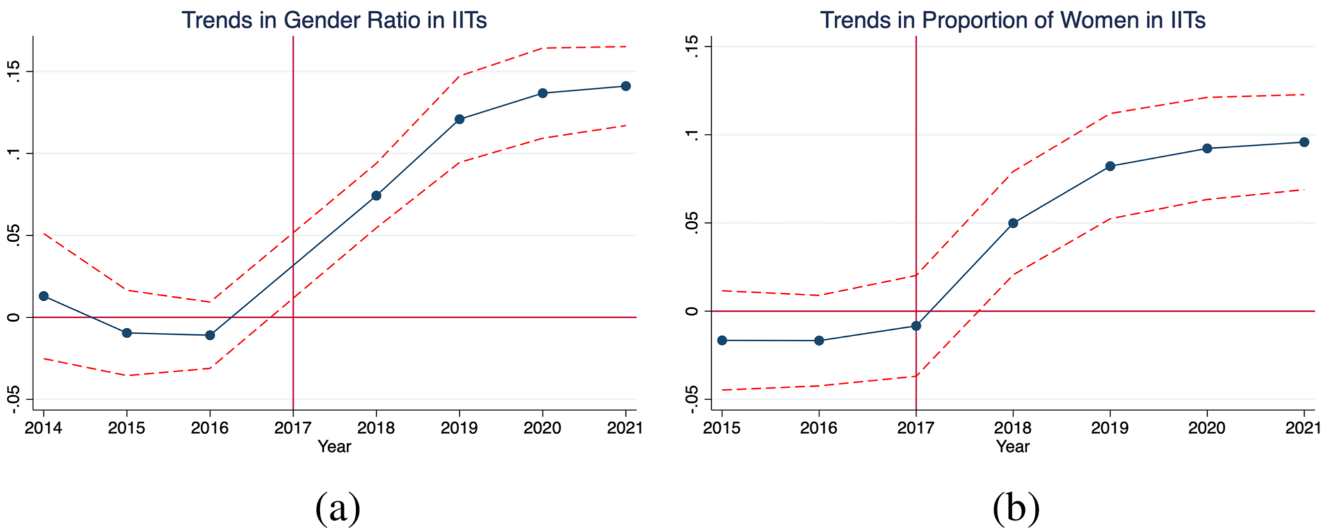


FIGURE 2 | Trends based on numbers from 20 IITs between 2014 and 2021. [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 3 | Location of 23 IIT Campuses. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 | Impact on female applications at IITs.

	Prop of women in registrations	Prop of women qualifying
Post	0.0192**	0.0182***
	(0.00651)	(0.00341)
Observations	49	49
Control mean	0.18	0.12
R^2	0.434	0.297
Number of IIT zones	7	7

Note: Data Source is IIT Entrance Exam Reports for years 2013–2020. The dependent variable is the total number of women who register for the IIT entrance exam (qualify the IIT entrance exam) divided by the total number of registrations (students who qualify the exam). The independent variable is a dummy taking the value 1 for post-policy years.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

general, that might not be true. It also offers all courses. There is, however, the social cost of traveling to college which depends on the distance (**d**) to the college from one's home and represents the social norms, safety concerns and long travel times. The framework is depicted in a decision tree in the picture below. The utility function and the trade offs faced by the girl in making her decisions are presented in the next subsections.

Definition 1. Girl lives **far** if $\gamma_E < \beta(d_E - d_L)$ and **close** otherwise.

3.2 | Decision at Stage 1

Case 1. Girl lives far from E.

As solved in Stage 2, she will never go to E if she chooses science track as the wage premium associated with E is not enough to cover for her cost of traveling. She will go to L with probability 1 if she chooses science irrespective of her admission outcome in E.

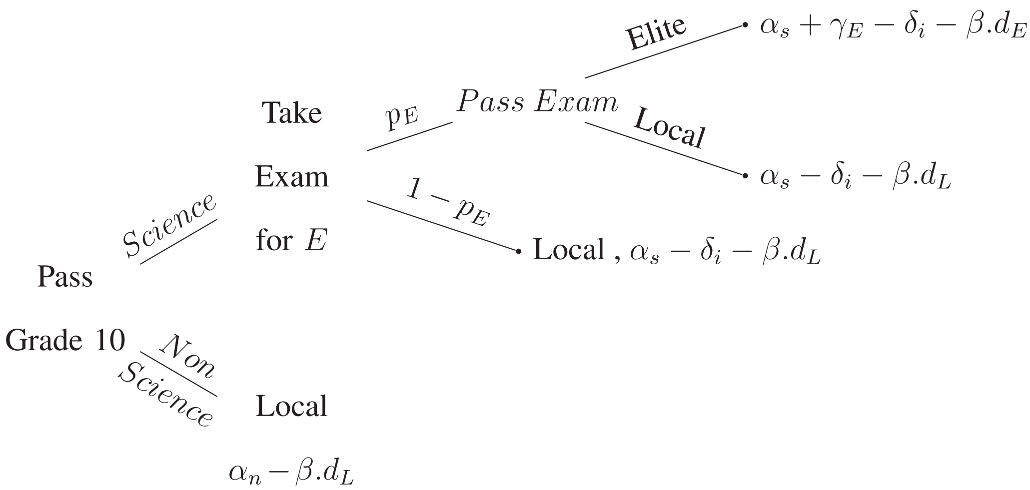
$$U^i = \begin{cases} \alpha_s - \delta_i - \beta.\, d_L & \text{if } S = 1 \\ \alpha_n - \beta.\, d_L & \text{if } S = 0 \end{cases}$$

She chooses science if the extra earnings from the science track are greater than the distaste associated with studying science.

$$\Rightarrow \delta_i < \alpha_s - \alpha_n \tag{3}$$

Case 2. Girl lives close to E.

As solved above, she will go to E if she chooses science track and gets admission into E (i.e., with probability p_E). She will go to L with probability $1 - p_E$ if she chooses science.



3.1 | Decision at Stage 2

Conditional on having chosen science and gotten admission into E, girl goes to E if the wage premium is greater than the cost associated with traveling the extra distance. Mathematically,

$$U_E > U_L$$

$$\Rightarrow \gamma_E > \beta(d_E - d_L) \tag{2}$$

Despite choice of science track and getting an admit, girl will not go to E (and go to L) if the above condition is not met.

$$U^i = \begin{cases} \alpha_s + p_E.\, \gamma_E - \delta_i - \beta\{p_E.\, d_E + (1 - p_E).\, d_L\} & \text{if } S = 1 \\ \alpha_n - \beta.\, d_L & \text{if } S = 0 \end{cases}$$

She chooses science if the extra earnings from the science track plus the expected increase in the wage premium associated with elite college net of the extra distance cost is greater than the distaste associated with studying science.

$$\Rightarrow \delta_i < (\alpha_s - \alpha_n) + p_E(\gamma_E - \beta(d_E - d_L)) \tag{4}$$

Proposition 1. *(a) If a girl lives close (*i.e., *Equation 2 is satisfied), it is optimal for her to choose science in school as long as her*

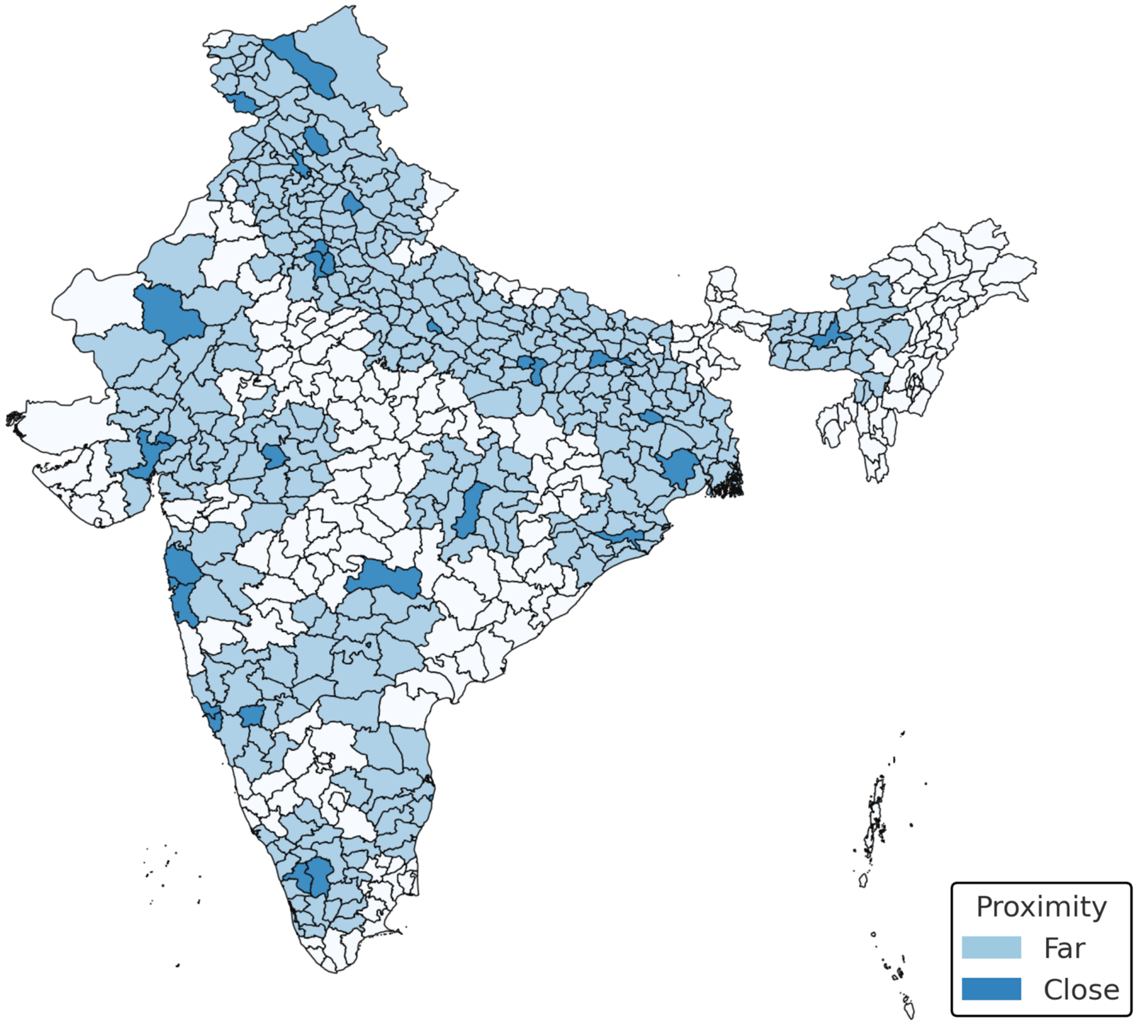


FIGURE 4 | Map showing treated (close) and control (far) districts. [Color figure can be viewed at wileyonlinelibrary.com]

*distaste for the subject is not too high (*i.e., *Equation 4 is satisfied). If she gets into the elite college, she studies a STEM course.*

(b) If a girl lives far (i.e., Equation 2 is not satisfied), then the choice of subject only depends on the distaste parameter (i.e., choose science if Equation 3 is satisfied). The decision is independent of the probability of getting into the elite college.

Corollary 1. *An affirmative action policy at an elite college will influence those girls who live close. Moreover, if the increase in probability of getting into the elite college is large enough to outweigh their distaste for STEM, they will switch to choosing science.*

4 | Data and Identification

4.1 | Data

I study the impact of this policy on subject choice by using the 75th round of the NSS that focuses on education. The survey was conducted between June 2017 and June 2018 and consists of a nationally representative sample of 64,519 rural households from 8097 villages and 49,238 urban households from 6188 blocks. The data cover qualitative and quantitative aspects of education such as educational attainment, access to schools and internet, educational expenditure and scholarships, type of education, and subject choice of individuals currently attending education. The policy was announced in April 2017, and I utilize this data to study the *announcement* effect of the policy by looking at the subject choice of young boys and girls below the age of 18 years who are being affected by the addition of supernumerary seats at elite engineering colleges across India. My analysis is restricted to a sample of individuals aged 13–24. Descriptive statistics are presented in Table 3.

4.2 | Identification

The main outcome of interest is the probability of studying science after Grade 10. The first difference compares this outcome between girls of age less than 18 years ("treated" cohort) who made their subject choice decisions after the policy was announced and older girls ("control" cohort) who would have already chosen their subject. For the second difference, boys are taken as the control group as they would have also been exposed to all other confounding factors such as a changing educational

TABLE 3 | Descriptive statistics.

	Young (Age ≤17)		Older (Age > 17)	
	Mean	**Standard deviation**	**Mean**	**Standard deviation**
Panel A: All				
Education level	8.33	1.25	11.26	1.60
Private coaching	0.29	0.45	0.16	0.37
Science	0.11	0.31	0.29	0.45
Panel B: Men				
Education level	8.33	1.25	11.17	1.52
Private coaching	0.30	0.46	0.16	0.37
Science	0.12	0.33	0.32	0.47
Panel C: Women				
Education level	8.32	1.25	11.38	1.70
Private coaching	0.27	0.44	0.16	0.37
Science	0.10	0.30	0.24	0.43

Note: The statistics are calculated for the individuals in the NSS Education Round 2017–2018.

TABLE 4 | Testing the parallel trends assumption.

Dependent variable: Probability of studying science		
Panel A: Parallel trends assumption for DID		
Age × Female	−0.0187***	
	(0.00330)	
Age	−0.0114***	
	(0.00242)	
Female	0.248***	
	(0.0673)	
Observations	29,105	
R^2	0.182	
Panel B: Parallel trends assumption for DDD		
Age × Female × Close	0.00127	0.000674
	(0.0101)	(0.0107)
Age × Female	−0.0186***	−0.0189***
	(0.00319)	(0.00406)
Close × Female	−0.0888	−0.0731
	(0.206)	(0.218)
Close × Age	0.00424	−0.00424
	(0.00723)	(0.00747)
Age	−0.0121***	−0.00432
	(0.00248)	(0.00294)
Female	0.256***	0.262***
	(0.0655)	(0.0838)
Close	0.115	0.533***
	(0.146)	(0.155)
Observations	29,105	28,805
Synthetic DID weights for districts	No	Yes
R^2	0.182	0.138

Note: This analysis uses individuals in the "control" cohort in NSS Education Round 2017–2018. I include district fixed effects, household-specific, and individual-level controls in the above regressions. Robust standard errors clustered at the district level are reported in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

environment and economic growth in the country, but the IITs only increased seats for girls. However, since the proportion of girls studying science is much lower than that of boys to begin with, it is plausible that the trends in the outcome for girls are different than those of boys. I therefore test for the parallel trends assumption for this double difference[8] in the pre-treatment cohorts. The coefficient on the interaction term in Table 4 Panel A is statistically significant, and therefore the null hypothesis of parallel trends is rejected.

In order to overcome the non-parallel trends between girls and boys, I conduct a triple difference analysis using proximity to an IIT campus as the exogenous source of variation. To study at an IIT, the decision to study science has to be taken before entering high school, that is, at a stage when most students are residing with their parents. Whether or not an IIT is close to a student's home is determined exogenously, and the place of residence is not affected by the location of an IIT or the introduction of the policy. This policy is introduced in a context where distance to home is a major determinant of educational and college choices, gender norms are prevalent, and crimes against women are rising.[9] These factors impede female mobility to access schools and colleges and limit their education choices. As presented in Table 6, distance to college matters, and it matters more for women; therefore, they travel to colleges closer to their homes. While most students at IITs live on campus, there is anecdotal evidence suggesting that girls tend to choose IITs closer to their home towns. A study by IIT Delhi faculty revealed that while a significant percentage of girls qualified for the JEE Advanced exam, only a smaller percentage actually joined IITs, with many opting for non-IITs (NITs and IIITs) to be closer to their homes and for branch preferences (Education Times 2023). Therefore, girls living closer to IITs have a comparative advantage in accessing such institutions over girls living far from IITs. I, therefore, define "close" ("treated") areas as those districts that lie within a 30 km radius of an IIT, and "far" ("control") areas are the ones that lie outside the 30 km radius but within a 200 km radius of an IIT.[10] I exclude the districts that are farther than 200 km from my analysis to reduce noise and improve precision. Moreover,

TABLE 5 | Difference-in-differences estimate.

Dependent variable: Probability of studying science	(1)
Young × Female	0.113***
	(0.00737)
Young	−0.188***
	(0.00896)
Female	−0.127***
	(0.00645)
Observations	59,664
Control Mean	0.24
R^2	0.187
District FE	Yes
Controls	Yes

Note: This regression compares the gender gap in the probability of studying science between older and younger cohorts. Robust Standard Errors clustered at the district level are reported in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

TABLE 6 | Proximity to educational institution for women.

Distance	(1)	(2)
Female	−0.243***	−0.163***
	(0.0109)	(0.0173)
Constant	3.003***	4.503***
	(0.0306)	(0.0293)
Observations	151,073	39,260
Sample	All	Above class 12
R^2	0.005	0.004

Note: The data used is the 75th round of National Sample Survey (2017–2018) dedicated to education. Dependent variable is a categorical variable for distance (d) of the educational institution from the place of residence for individuals currently attending education. It is coded as: 1 for $d < 1$ km, 2 for $1\,\text{km} < d < 2\,\text{km}$, 3 for $2\,\text{km} < d < 3\,\text{km}$, 4 for $3\,\text{km} < d < 5\,\text{km}$ and 5 for $d > 5\,\text{km}$.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

districts that are too far can be very different from districts closer to IITs. The triple difference estimate is constructed by taking the difference between the double difference in close districts with that of the far districts. My identifying assumption is that conditional on district-specific characteristics and individual-level controls, gender gaps in science across closer and farther districts from IIT would be parallel across different age groups in the absence of the policy. If the parallel trends assumption is satisfied, the triple difference will causally estimate the change in probability of choosing science subjects in high school. I test the identifying assumption in Table 4 Panel B in the pre-treatment cohort, and I cannot reject the null hypothesis of parallel trends.

4.2.1 | Synthetic Difference-in-Differences

A potential issue with using distance to IIT is that districts close to IITs can be quite different from districts that are far and therefore can probably not be considered a good control.[11] I control for individual and household-specific characteristics in my model and include district fixed effects to capture any time-invariant differences across these districts. Moreover, district-specific differences are common for boys and girls and will get canceled out with the triple difference. Since districts that are close to IITs versus those that are farther may evolve differently over time, I further assign SDIDs weights (Arkhangelsky et al. 2021) to control districts. This approach is an advanced version of the synthetic control method (Abadie et al. 2010) which is used in panel datasets to correct for parallel trends by assigning unit-specific and time-specific weights. Since I have a cross-section, I first categorize the data by the age category of the individual to create a time dimension that is individuals below the age of 18 are young, between 18 to 22 are middle, and those with age higher than 22 are categorized old. Then, I collapse the data at the age category and district level to run a SDID regression for the main outcome.[12] The method assigns synthetic weights to control districts (those that are farther than 30 km) and to the pre-treatment time periods (the "middle" and "old" cohort in this case) in order to obtain balance between close and far districts in each of the pre-treatment periods. Using unit-specific weights, I again fail to reject the null of parallel trends for triple difference as shown in Table 4 Panel B Column 2.[13]

5 | Estimating Equation and Results

I estimate the triple difference estimate in the following manner:

$$\begin{aligned} y_{iaj} = \alpha + \delta Female_{ij}.\ Young_i.\ Close_j + \beta_1 Female_{ij}.\ Young_i \\ + \beta_2 Close_j.\ Female_{ij} + \ \beta_3 Young_i.\ Close_j + \beta_4 Young_i \\ + \beta_5 Female_{ij} + \beta_6 Close_j + \rho X_i + \mu_s + \epsilon_{iaj} \end{aligned} \tag{5}$$

where y_{iaj} is an outcome variable of individual i of age a living in district j, $Young_i$ is takes value 1 if individual i's age a is less than 18 (i.e., the treated cohort), $Female_{iaj}$ is a dummy that takes value 1 if i is a female, $Close_j$ is a dummy that takes value 1 if district j lies within a radius of 30 km of an IIT & 0 if district j lies within a radius of 200 km of an IIT but farther than 30 km, μ_j represent the state (or district) fixed effects and X_i denote individual specific controls such as religion, caste, household consumption expenditure, whether household owns a computer and whether household owns an internet facility. I report robust standard errors clustered at the district level. The parameter of interest, δ, provides the triple difference (DDD) estimate of the change in probability of choosing science amongst girls.

I first estimate Equation (5) for the main outcome of interest—likelihood of studying science after Grade 10. The dependent variable is a dummy that takes value 1 if the individual reports choosing Science or Engineering as their discipline after Grade 10, and 0 otherwise. Table 7 Column 1 provides a triple difference estimate when I choose districts farther than 30 km as the far districts. I observe a 6.7 percentage point increase in the likelihood of choosing science track amongst girls. In column 2, I use SDID weights and the estimate increases to 7.4. Compared to the baseline mean, these results imply that since the knowledge about implementation of this policy has come into the public domain, girls are 27%–30% more likely to choose science as their subject after passing Grade 10, possibly because they anticipate that the choice of this subject is now associated with a higher probability of admission at a reputed engineering college.

In order to look at the impact on younger girls, I do two things. First, I repeat my analysis by changing the age cutoff that I use to determine the treatment cohort. In my main results, younger cohorts are the ones whose age is below 18. I also conduct a triple difference analysis by re-defining the treated cohort as the ones below age a where $a \in \{15,16,17,19,20,21,22\}$. The triple difference coefficients are plotted in Figure 5. As I change the treated cohort, the effect diminishes for higher age cutoffs. The effect is still significant when younger ages (below 18) are used as cutoff which shows that the policy affected younger girls. However, the effect is less precise amongst lower ages since the subject choice is made at the higher secondary school level and hence we see the most effect amongst students who are closest to making their decision when the policy was announced. Second, I conduct an event study regression where I interact different age dummies with female and close dummy. This provides age-specific coefficients of the triple interaction as plotted in Figure 6. The numbers on the x-axis represent how smaller the age is from the cutoff age of 18. Therefore, the coefficients on the right of 0 represent younger individuals. This further shows evidence for parallel trends in the older cohorts as the coefficients to the left of 0 are insignificant. This shows that gender gaps in science moved in parallel between treated and control districts in the older cohorts. Positive coefficients on the right highlight the finding that younger females exposed to the policy in the treated districts are more likely to study science.

TABLE 7 | Triple difference analysis.

Dependent variable: Probability of choosing science	**(1)**	**(2)**
Young × Female × Close	0.0665***	0.0740***
	(0.0235)	(0.0254)
Female × Young	0.103***	0.0888***
	(0.00767)	(0.0111)
Close × Female	−0.0682***	−0.0802***
	(0.0209)	(0.0238)
Young × Close	−0.0337	−0.0527*
	(0.0214)	(0.0311)
Young	−0.183***	−0.170***
	(0.00911)	(0.0242)
Female	−0.116***	−0.0980***
	(0.00682)	(0.0119)
Close	0.0958***	0.281***
	(0.0187)	(0.0260)
Control mean	0.25	0.23
Synthetic DID weights	No	Yes
Observations	59,664	58,592
R^2	0.188	0.159
District FE	Yes	Yes

Note: This is the main regression. Younger females living closer to IITs are more likely to study science. Robust Standard Errors clustered at the district level are reported in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

5.1 | Other Outcomes

I also estimate Equation (5) using the highest level of education attained as the dependent variable. Such policies are meant to encourage higher education in general for girls and can have a positive impact. However, since this policy was implemented in elite institutions where students have to face very aggressive competition to enter and therefore specifically targets girls with high ability, the effect on educational attainment can be negligible as compared to the population as a whole. I test this hypothesis and report the triple difference in Table 8. I do not find any evidence of a differential impact on the educational attainment of girls living in areas closer to IITs. This suggests the absence of any other educational program or intervention that could be in place to differently impact girls' educational attainment, and any impact on subject choice should be coming from the supernumerary policy.

I also look at the impact on uptake of private coaching amongst younger girls. The preparation for qualifying the IIT entrance examinations often involves not studying Science subjects but also requires rigorous training for qualifying the competitive examinations. Therefore, students indulge in private coaching or tuition through established coaching centers which aim towards that. The triple difference estimate for this outcome is reported in Table 9. I do not find a statistically significant increase in the likelihood of taking private coaching amongst younger girls living close to elite colleges. While it is possible that the policy did not change the private coaching uptake, we should look at this result with caution. In the data that I use, individuals are asked whether they currently take private coaching or not but do not specify if they took private coaching in higher secondary classes. It is possible that students in all age cohorts have joined private coaching at some point and for different reasons, which makes it difficult to disentangle whether the private coaching was to prepare for elite engineering colleges or whether it was for something else and specifically when the private coaching was taken.

Finally, I look at some expenditure outcomes where data is collected on the total expenditure made on studying a basic

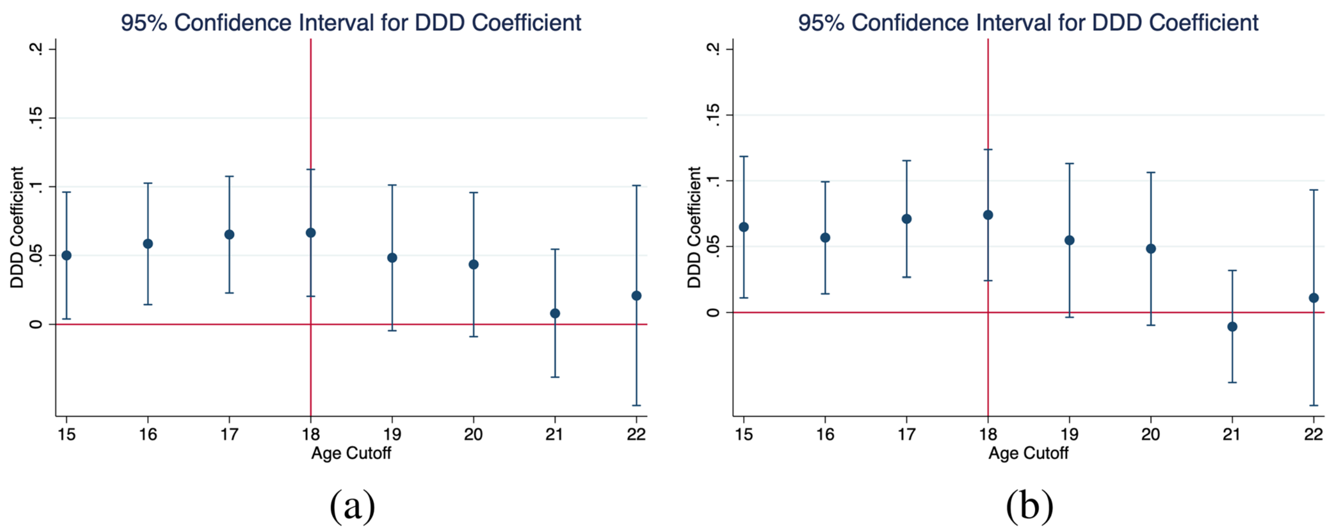


FIGURE 5 | Impact on younger girls—coefficients using different ages as treatment. (a) Without SDID weights. (b) With SDID weights. [Color figure can be viewed at wileyonlinelibrary.com]

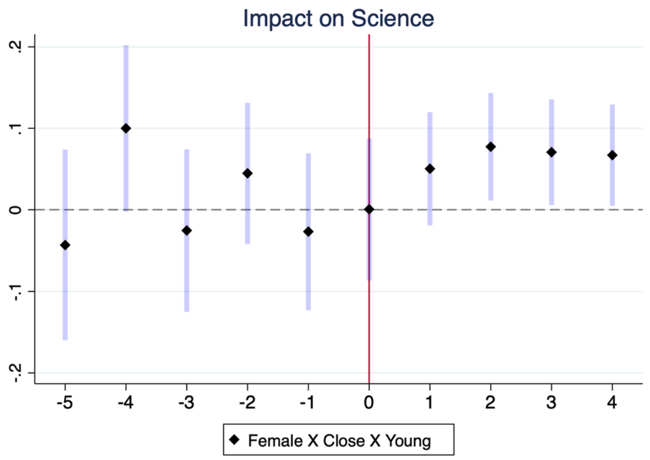


FIGURE 6 | Event study: Impact on younger (post-policy) girls (90% CIs). [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 8 | Educational attainment.

Dependent variable: Educational level	(1)	(2)
Young × Female × Close	−0.0414	−0.2643
	(0.0813)	(0.2069)
Synthetic DID weights	No	Yes
Observations	59,664	58,592
R^2	0.529	0.508
District FE	Yes	Yes

Note: The regression shows no impact on educational attainment. Robust Standard Errors clustered at the district level are reported in parentheses.
$^{*}p<0.10$.
$^{**}p<0.05$.
$^{***}p<0.01$.

TABLE 9 | Uptake of private coaching.

Dependent variable: Private coaching uptake	(1)	(2)
Young × Female × Close	0.0208	0.00599
	(0.0245)	(0.0283)
Synthetic DID weights	No	Yes
Observations	59,664	58,592
R^2	0.257	0.217
District FE	Yes	Yes

Note: The regression shows no impact on private coaching uptake. Robust Standard Errors clustered at the district level are reported in parentheses.
$^{*}p<0.10$.
$^{**}p<0.05$.
$^{***}p<0.01$.

course in the current academic year, expenditure on extra tuition, and expenditure on preparation for higher studies. The reported expenditure is in Rupees. Again, I do not find a significant effect on any of these outcomes, as shown in Table 10. Even though the policy seems to have pushed the choices of girls towards studying science, the amount of expenditure incurred for their education does not seem to be changing. However, I would look at these results with caution, as well as the data collected, which asks about expenditure in the current academic year and not at a particular time of the individual's life.

6 | Possible Mechanisms

I explore possible mechanisms that can drive the subject choice decisions among girls as a consequence of this policy.

6.1 | Sibling Spillover Effect

It is well documented that subject choice is at times influenced by the decision made by elder siblings. There is an incentive to choose a subject when the elder sibling has studied the same subject so as

TABLE 10 | Other expenditure outcomes (in INR).

	Log total expenditure	Log total expenditure	Tuition	Tuition	Higher studies prep	Higher studies prep
Young × Female × Close	−0.0427	−0.340	237.6	110.5	50.70	50.60
	(0.0527)	(0.278)	(214.5)	(256.5)	(73.60)	(74.06)
SDID weights	No	Yes	No	Yes	No	Yes
Observations	44,939	43,908	59,630	58,558	59,630	58,558
R^2	0.344	0.324	0.165	0.137	0.022	0.018
District FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The regression shows no impact on other parental investments. Robust Standard Errors clustered at the district level are reported in parentheses.
$^{*}p < 0.10$.
$^{**}p < 0.05$.
$^{***}p < 0.01$.

TABLE 11 | Sibling spillover effects.

Dependent variable: Probability of choosing science	(1)	(2)
Young × Close × Female × Sibling science	−0.101	−0.114
	(0.0841)	(0.0827)
Young × Close × Female	0.0728***	0.0788***
	(0.0217)	(0.0233)
Young × Close × Sibling science	0.231***	0.261***
	(0.0586)	(0.0640)
Sibling science	0.121***	0.136***
	(0.0245)	(0.0300)
SDID weights	No	Yes
Observations	59,664	58,592
R^2	0.189	0.160
District FE	Yes	Yes

Note: Robust Standard Errors clustered at the district level are reported in parentheses.
$^{*}p < 0.10$.
$^{**}p < 0.05$.
$^{***}p < 0.01$.

to benefit from their resources, experience and knowledge. While elder sibling can encourage an individual to take the same subject, it is also possible that elder sibling's (bad) experience can deter an individual from studying the same subject. I explore whether the increase in likelihood of studying science after the introduction of the policy is driven amongst girls whose elder sibling also studied science. I define a variable "Sibling Science" which takes value 1 if individual i has an elder sibling who studied Science and 0 otherwise. I interact this variable with the triple difference to determine heterogeneity. The fourth difference is insignificant as shown in Table 11. Younger students living closer to IITs are more likely to study Science if their elder sibling also studied Science but this effect is not significantly different for girls. The effect of the policy does not seem to be driven differently amongst girls who have an elder sibling who also studied Science.

TABLE 12 | Heterogeneity by parent's education level.

Dependent variable: Probability of choosing science	(1)	(2)
Young × Close × Female × Parent education	−0.000648	0.00399
	(0.00908)	(0.0115)
SDID weights	No	Yes
Observations	55,863	54,830
R^2	0.195	0.168
District FE	Yes	Yes

Note: Robust Standard Errors clustered at the district level are reported in parentheses.
$^{*}p < 0.10$.
$^{**}p < 0.05$.
$^{***}p < 0.01$.

6.2 | Parents' Education

I also explore heterogeneity by parents' education level. The results are presented in Table 12. I first identify parents of each individual in the data. I then determine the highest level of education obtained by each parent. I take the maximum of the educational attainment of the two parents. The parent education variable is a continuous variable which determines the educational level attainment of the more educated parent. I interact this variable with the triple difference to determine the heterogeneity. I do not find heterogeneous effects by the parents' education level, that is, the increase in likelihood of studying science amongst younger girls living closer to IITs does not increase as parents' education level rises.

7 | Robustness and Sensitivity Checks

7.1 | Excluding the Old IITs

Out of the 23 IITs, 7 IITs were established between 1951 and 1963, because of which they continue to be top-ranked owing to their renowned curriculum, faculty, infrastructure, and job market placements. As a robustness check, I remove these 7 IITs

TABLE 13 | Triple difference analysis for 16 new IITs.

Dependent variable: Probability of choosing science	(1)	(2)
Young × Female × Close	0.0971***	0.103***
	(0.0319)	(0.0380)
Observations	35,464	34,494
Control mean	0.22	0.24
Synthetic DID weights	No	Yes
R^2	0.183	0.156
District FE	Yes	Yes

Note: Robust Standard Errors clustered at the district level are reported in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

TABLE 14 | Restricting the sample.

Dependent variable: Probability of choosing science	(1)	(2)	(3)
Young × Female × Close	0.0726**	0.0653**	0.0634**
	(0.0299)	(0.0268)	(0.0259)
Observations	14,839	24,029	36,244
R^2	0.192	0.181	0.179
Band	30 km	30 km	30 km
Far control	60 km	90 km	120 km
District FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Note: Robust Standard Errors clustered at the district level are reported in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

TABLE 15 | Distance threshold.

Dependent variable: Probability of choosing science	(1)	(2)	(3)
Young × Female × Close	0.0795***	0.0489**	0.0379*
	(0.0251)	(0.0213)	(0.0205)
Observations	59,664	59,664	59,664
R^2	0.188	0.188	0.187
Band	20 km	40 km	50 km
District FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Note: Robust Standard Errors clustered at the district level are reported in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

TABLE 16 | Dyadic comparison: Average DDD coefficient.

Probability of choosing science	(1)	(2)
Young × Female × Close	0.0672***	0.0606***
	(0.0156)	(0.01504)
Synthetic DID weights	No	Yes

Note: Robust Standard Errors clustered at the district level are reported in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

TABLE 17 | IIT zone fixed effect.

Dependent variable: Probability of choosing science	(1)	(2)
Young × Close × Female	0.0665***	0.0906***
	(0.0241)	(0.0311)
Observations	59,664	58,592
Control mean	0.25	0.23
Synthetic DID weights	No	Yes
R^2	0.117	0.140
Band	30 km	30 km
IIT zone FE	Yes	Yes
Controls	Yes	Yes

Note: Robust Standard Errors clustered at the district level are reported in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

TABLE 18 | Including district by rural/urban FE.

Dependent variable: Probability of choosing science	(1)	(2)
Young × Female × Close	0.0635***	0.0727***
	(0.0235)	(0.0254)
Observations	59,664	58,592
Control mean	0.25	0.23
Synthetic DID weights	No	Yes
R^2	0.197	0.162
District-rural/Urban FE	Yes	Yes

Note: Robust Standard Errors clustered at the district level are reported in parentheses.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

from my analysis to check if results are driven by these popular IITs. I present the triple difference estimate in Table 13 and find that there is an increase in the likelihood of choosing the science track even if we only consider the relatively new IITs. The estimate is larger in magnitude, indicating that the effect is being driven in areas with newer IITs.

TABLE 19 | Sensitivity check: Dropped one IIT a time.

IIT dropped	DDD estimate	DDD estimate using SDID weights
(BHU) Varanasi	0.0574**	0.0708***
	(0.0230)	(0.0260)
(ISM) Dhanbad	0.0637**	0.0766***
	(0.0245)	(0.0270)
Bhilai	0.0615**	0.0755***
	(0.0243)	(0.0268)
Bhubaneshwar	0.0535**	0.0685**
	(0.0267)	(0.0280)
Mumbai	0.0619**	0.0794***
	(0.0249)	(0.0268)
Delhi	0.0584**	0.0589**
	(0.0259)	(0.0255)
Dharwad	0.0628**	0.0760***
	(0.0244)	(0.0280)
Gandhinagar	0.0605**	0.0756***
	(0.0245)	(0.0265)
Goa	0.0596**	0.0750***
	(0.0241)	(0.0268)
Guwahati	0.0615**	0.0768***
	(0.0246)	(0.0268)
Hyderabad	0.0691***	0.0829***
	(0.0253)	(0.0275)
Indore	0.0747***	0.0875***
	(0.0243)	(0.0276)
Jammu	0.0621**	0.0764***
	(0.0243)	(0.0269)
Jodhpur	0.0562**	0.0723**
	(0.0250)	(0.0276)
Kanpur	0.0667***	0.0793***
	(0.0240)	(0.0269)
Kharagpur	0.0673***	0.0821***
	(0.0251)	(0.0275)
Madras	0.0552**	0.0693**
	(0.0266)	(0.0288)
Mandi	0.0628**	0.0763***
	(0.0249)	(0.0271)
Palakkad	0.0756***	0.0898***
	(0.0254)	(0.0265)

(Continues)

TABLE 19 | (Continued)

IIT dropped	DDD estimate	DDD estimate using SDID weights
Patna	0.0616**	0.0760***
	(0.0242)	(0.0266)
Roorkee	0.0653***	0.0762***
	(0.0239)	(0.0273)
Ropar	0.0622**	0.0769***
	(0.0245)	(0.0267)
Tirupati	0.0568**	0.0734***
	(0.0246)	(0.0272)

Note: Robust Standard Errors clustered at the district level are reported in parentheses.
$^{*}p < 0.10$.
$^{**}p < 0.05$.
$^{***}p < 0.01$.

7.2 | Redefining Far Districts

The main analysis is restricted to districts which are atmost 200 km away from an IIT. I adjust my sample size by considering districts which are atmost 60, 90, and 120 km away from an IIT. This reduces the sample size at my disposal. I compare the gender gap in the likelihood of choosing science between districts less than 30 km away and those that are farther. I still observe robust estimates of the policy impact as shown in Table 14.

7.3 | Changing the Distance Threshold

I test the sensitivity of my results to the threshold level that differentiates a close and a far district in Table 15. I repeat the analysis when a district is close if within 20, 40, or 50 km of an IIT. I find that the impact of the policy is higher when the treated district is within 20 km but fades away as the treated district gets farther from an IIT. The policy, therefore, affects the most who live close to the elite colleges (as concluded in Section 3).

7.4 | Dyadic Comparisons

The main analysis compares districts close to any IIT with those of districts far from any IIT. Additionally, I compare the gender gap in science across cohorts between close and far districts *within each IIT zone*. I first divide the districts in the data into 23 IIT zones depending on which IIT is closest to that district. For instance, an IIT-Delhi zone consists of all those districts for which the closest IIT is IIT Delhi. For each separate IIT zone, I perform the usual triple difference regression, which compares the gender gap in the likelihood of choosing science before and after the policy between the districts within a 30 km radius with those that are outside that radius but within the same zone. I report the average triple difference coefficient from the regressions of 20 IIT zones[14] in Table 16. The average coefficient of the individual regressions

TABLE 20 | Balance table for comparing far and close districts.

Variable	Control	Treated	Diff	*p*
Household expenditure	11468.42	14943.95	−3475.53***	0.000
Age	17.17	17.45	−0.28**	0.028
Education level	9.84	10.09	−0.25**	0.011
Proportion of females	0.41	0.40	0.01	0.573
Household computer ownership	0.21	0.40	−0.19***	0.000
Household size	5.34	5.08	0.26*	0.059
Household internet usage	0.61	0.78	−0.17***	0.000
Probability of studying science	0.20	0.23	−0.03	0.203
Joined private tutoring	0.18	0.20	−0.02	0.584

Note: The table compares characteristics of far and close districts from the NSS Education Survey 2017–2018.
*$p < 0.10$.
**$p < 0.05$.
***$p < 0.01$.

pertaining to each IIT zone is positive, statistically significant, and similar in magnitude to the triple difference coefficient that I obtained in Table 7.

7.5 | Including IIT-Zone and District by Region Fixed Effects

As an additional robustness check, I include IIT zone fixed effects in the main results. The results are robust when I control for any time-invariant IIT-zone specific characteristics (Table 17). The use of synthetic DID weights increases the average treatment effect. The results are also robust when I include district by rural–urban fixed effects (Table 18).

7.6 | Dropping One IIT at a Time

The results are robust when I drop one IIT at a time from the regression, as shown in Table 19. The point estimates vary between 0.053 and 0.076. They vary between 0.058 and 0.089 when SDID weights are used.

8 | Discussion

In this paper, I show that reserving seats for girls in elite STEM colleges can impact subject choices in school. While we do see increases in enrolment at these elite STEM colleges, the increase in seats is probably insufficient to meet the increasing demand for STEM courses amongst girls, especially if all the girls who switch to study Science subjects in school actually do pursue engineering courses. That is to say, there is possibly an impact on enrolment in other elite or non-elite colleges as well where this policy was not introduced. There are possible spillover effects of this policy on other institutions. Due to the paucity of data and because there are a large number of such institutions, I am unable to estimate the spillover effect in this paper. If such effects are there and if they are non-negative, the policy can have a much larger impact on undergraduate STEM enrolment in the country as a whole. I use distance to IIT as the identifying channel in order to define my treatment. Another possible channel that is correlated with distance is the information channel. It is important to note that the information about the policy is more relevant in areas closer to IITs than those that are farther. This is especially true since areas closer to IITs often form hubs which provide training and coaching for clearing their qualifying competitive exams. Therefore, any policy-related information that especially pertains to IITs is expected to spread through these hubs which are more likely to be closer to IITs than farther from them.

A possible concern that can be raised given the identification strategy that I use is that students relocate to areas closer to IITs or other places in the country which provide extensive training for the competitive exam and stay away from their parents after passing secondary schooling. However, the sample of my study consists of individuals living with their families and those students are not captured in the data that I use. Therefore, if any girls are moving out of the city in order to attend coaching as they anticipate improved chances of their admission, they would not be captured in my data and I would be underestimating the effect of the policy. However, I do not think the policy would make individuals switch subjects as well as make girls migrate out of hometown at the same time. The girls who believe that the policy has changed their probability of admission into IITs to one would be pursuing science anyway and if they were to make the choice would migrate for better coaching opportunities anyway. In a context where social norms are salient, a joint decision of studying science as well as migrating out of hometown for the same comes at a huge cost, whereas the benefits are only marginal.

Lastly, I would like to mention a possible mechanism which can be driving the effect of the policy—peer effects. Peers play a pivotal role in influencing school, college, and subject decisions, especially within the same gender in developing areas with strict gender norms. The information about the policy is more likely to spread in peer groups. As girls discover the policy, they are more likely to inform their friends about it. They also tend to choose the same subjects as their friends so that they can spend more time together, stay in the same class, or join the same tuition. Therefore, one would expect the policy to have a larger effect when their peers also choose Science. Due to insufficient data on the subject choice of peer groups or social networks, I cannot explore this mechanism and whether such networks and peer effects are driving the results is an open question for future research.

Acknowledgments

This paper has benefited from discussions with Sheetal Sekhri, Amalia Miller, Gaurav Chiplunkar, and participants at University of Virginia Development Workshops, IIT-Delhi Econ Talk Seminar 2021, and ISI Delhi Growth Conference 2023. This paper was awarded the Best Summer Paper Award 2021 by the Dept of Economics at UVA and the Huskey Research Prize 2023 by the UVA GSAS Council.

Funding

The author has nothing to report.

Conflicts of Interest

The author declares no conflicts of interest.

Endnotes

[1] Countries like France, Germany, etc. also impose a first-level subject choice at the school level.

[2] Statista survey shows that about 70% urban dwellers across India traveled less than 10 km and spent around 27 min on average to travel for work and education in 2019.

[3] A map of all 23 IIT campuses is provided in Figure 3.

[4] The course offers specialization in various engineering fields such as Computer Science, Electrical, Electronics and Communication, Information Technology etc.

[5] The increase in girls is statistically significant at 1% but is insignificant for boys.

[6] The criteria for qualifying the exam is based on cut-off score in the entrance exam which could be directly proportional to the number of seats IIT added to increase the gender ratio.

[7] Because of this assumption, the model cannot comment on possible spillovers of any policy at the elite college on other colleges.

[8] The DID estimate corresponding to the double difference is reported in Table 5.

[9] In 2019, cases registered under crime against women rose by 7% relative to 2018. As per the National Crime Records Bureau (NCRB) report 2020, an average of 87 rape cases were registered daily in India in 2019.

[10] Figure 4 presents a map showing the close ("treated") and far ("control") districts. While there is no direct evidence for the choice of 30 km threshold, it is a reasonable commuting distance. Commuting daily is common in India as 70% urban dwellers spend 27 min on average to travel less than 10 km for work and education purposes (Statista 2020). The median distance traveled for college in the United States is 17 miles (or about 27 km) (Hillman 2023). With these estimates, a rough estimate for a commutable distance in India is 30 km. This threshold is not too high to inhibit an average individual to commute to college daily or for parents to make visits, as well as not too low that would reduce the power in my analysis. Moreover, IITs allow individuals to choose not to stay at university dorms as long as they are within that radius.

[11] Table 20 shows differences between close and far districts. I control for these characteristics in my regression. The districts are balanced on the likelihood of studying science—main outcome of interest.

[12] Some districts are dropped in the analysis to make sure that district-age cohort panel is strongly balanced for the SDID to work.

[13] The assumption of parallel trends is not violated even if I use both district and age specific weights but I only show the result with district-specific weights as this regression only includes the pre-treatment data.

[14] Zones of IIT Bhilai, IIT Ropar and IIT Jammu are omitted due to lack of sufficient data.

References

Abadie, A., A. Diamond, and J. Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105, no. 490: 493–505.

Adams, R. B., and T. Kirchmaier. 2016. "Women on Boards in Finance and STEM Industries." *American Economic Review* 106, no. 5: 277–281.

Altmejd, A., A. Barrios Fernández, M. Drlje, et al. 2021. "O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries." *Quarterly Journal of Economics* 136, no. 3: 1831–1886.

Arcidiacono, P., E. M. Aucejo, and V. J. Hotz. 2016. "University Differences in the Graduation of Minorities in STEM Fields: Evidence From California." *American Economic Review* 106, no. 3: 525–562.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. 2021. "Synthetic Difference-in-Differences." *American Economic Review* 111, no. 12: 4088–4118.

Aygün, O., and B. Turhan. 2017. "Large-Scale Affirmative Action in School Choice: Admissions to IITs in India." *American Economic Review* 107, no. 5: 210–213.

Backes, B. 2012. "Do Affirmative Action Bans Lower Minority College Enrollment and Attainment?: Evidence From Statewide Bans." *Journal of Human Resources* 47, no. 2: 435–455.

Bagde, S., D. Epple, and L. Taylor. 2016. "Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India." *American Economic Review* 106, no. 6: 1495–1521.

Bastarrica, M. C., N. Hitschfeld, M. M. Samary, and J. Simmonds. 2018. "Affirmative Action for Attracting Women to STEM in Chile." In *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, 45–48. Association for Computing Machinery (ACM).

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics* 124, no. 4: 1497–1540.

Beede, D. N., T. A. Julian, D. Langdon, G. McKittrick, B. Khan, and M. E. Doms. 2011. "Women in STEM: A Gender Gap to Innovation." In *Economics and Statistics Administration Issue Brief*, 4–11. U.S. Department of Commerce, Economics and Statistics Administration (ESA).

Bertrand, M., R. Hanna, and S. Mullainathan. 2010. "Affirmative Action in Education: Evidence From Engineering College Admissions in India." *Journal of Public Economics* 94, no. 1–2: 16–29.

Bleemer, Z. 2022. "Affirmative Action, Mismatch, and Economic Mobility After California's Proposition 209." *Quarterly Journal of Economics* 137, no. 1: 115–160.

Borker, G. 2017. "Safety First: Perceived Risk of Street Harassment and Educational Choices of Women." In *Job Market Paper, Department of Economics*, 12–45. Brown University.

Buser, T., M. Niederle, and H. Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *Quarterly Journal of Economics* 129, no. 3: 1409–1447.

Buser, T., N. Peter, and S. C. Wolter. 2017. "Gender, Competitiveness, and Study Choices in High School: Evidence From Switzerland." *American Economic Review* 107, no. 5: 125–130.

Card, D., and A. A. Payne. 2021. "High School Choices and the Gender Gap in STEM." *Economic Inquiry* 59, no. 1: 9–28.

Ceci, S. J., and W. M. Williams. 2015. "Women Have Substantial Advantage in STEM Faculty Hiring, Except When Competing Against More-Accomplished Men." *Frontiers in Psychology* 6: 1532.

Chanana, K. 2007. “Globalisation, Higher Education and Gender: Changing Subject Choices of Indian Women Students.” *Economic and Political Weekly* 27, no. 7: 590–598.

Daymont, T. N., and P. J. Andrisani. 1984. “Job Preferences, College Major, and the Gender Gap in Earnings.” *Journal of Human Resources* 19: 408–428.

Deem, R. 2012. *Women & Schooling*. Routledge.

Education Times. 2023. “Thanks to Supernumerary Quota, Girl Power Rules JEE Advanced.” Accessed June 11, 2025. https://www.educationtimes.com/article/campus-beat-collegelife/99735820/thanks-to-supernumerary-quota-girl-powerrulesjeeadvanced.

Fischer, S. 2017. “The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence.” *Labour Economics* 46: 211–226.

Fryer, R. G., Jr., and S. D. Levitt. 2010. “An Empirical Analysis of the Gender Gap in Mathematics.” *American Economic Journal: Applied Economics* 2, no. 2: 210–240.

Gupta, N. 2020. “Patriarchy Reinforced or Challenged? A Study of Engineering Students in an Elite Indian Institute.” *Gender, Technology and Development* 24, no. 2: 250–270.

Hillman, N. 2023. *Geography of Opportunity*. Institute for College Access & Success (TICAS). https://ticas.org/wp-content/uploads/2023/11/HIllman-Geography-of-Opportunity-Brief-2_2023.pdf.

Hinrichs, P. 2012. “The Effects of Affirmative Action Bans on College Enrollment, Educational Attainment, and the Demographic Composition of Universities.” *Review of Economics and Statistics* 94, no. 3: 712–722.

Jain, T., A. Mukhopadhyay, N. Prakash, and R. Rakesh. 2022. “Science Education and Labor Market Outcomes in a Developing Economy.” *Economic Inquiry* 60, no. 2: 741–763. https://doi.org/10.1111/ecin.13044.

Khanna, G. 2020. “Does Affirmative Action Incentivize Schooling? Evidence From India.” *Review of Economics and Statistics* 102, no. 2: 219–233.

Maceira, H. M. 2017. “Economic Benefits of Gender Equality in the EU.” *Intereconomics* 52, no. 3: 178–183.

Matheson, K., A. Echenberg, D. M. Taylor, D. Rivers, and I. Chow. 1994. “Women's Attitudes Toward Affirmative Action: Putting Actions in Context 1.” *Journal of Applied Social Psychology* 24, no. 23: 2075–2096.

Matsa, D. A., and A. R. Miller. 2013. “A Female Style in Corporate Leadership? Evidence From Quotas.” *American Economic Journal: Applied Economics* 5, no. 3: 136–169.

Miller, A. R., and C. Segal. 2019. “Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence.” *Review of Economic Studies* 86, no. 5: 2220–2247.

Munoz-Boudet, A. M., and A. Revenga. 2017. *Breaking the STEM Ceiling for Girls*. Brookings.

Niederle, M., and L. Vesterlund. 2010. “Explaining the Gender Gap in Math Test Scores: The Role of Competition.” *Journal of Economic Perspectives* 24, no. 2: 129–144.

Nollenberger, N., N'u. Rodrıguez-Planas, and A. Sevilla. 2016. “The Math Gender Gap: The Role of Culture.” *American Economic Review* 106, no. 5: 257–261.

Porter, C., and D. Serra. 2020. “Gender Differences in the Choice of Major: The Importance of Female Role Models.” *American Economic Journal: Applied Economics* 12, no. 3: 226–254.

Resmini, M. 2016. “The ‘Leaky Pipeline’.”

Sahoo, S., and S. Klasen. 2021. “Gender Segregation in Education: Evidence From Higher Secondary Stream Choice in India.” *Demography* 58, no. 3: 987–1010.

Sekhri, S. 2020. “Prestige Matters: Wage Premium and Value Addition in Elite Colleges.” *American Economic Journal: Applied Economics* 12, no. 3: 207–225.

Sharpe, S. 1976. “Just Like A Girl.” In *How Girls Learn to be Women*. Penguin Books.

Sloane, C., E. Hurst, and D. Black. 2019. *A Cross-Cohort Analysis of Human Capital Specialization and the College Gender Wage Gap*. National Bureau of Economic Research.

Speer, J. D. 2023. “Bye Bye Ms. American Sci: Women and the Leaky STEM Pipeline.” *Economics of Education Review* 93: 102371.

Statista. 2020. “Daily Commuting Distance Among Urban Dwellers in India as of February 2020.” Accessed June 11, 2025. https://www.statista.com/statistics/1103900/india-urbandwellers-daily-commutedistance/.

Sukhtankar, S., G. Kruks-Wisner, and A. Mangla. 2022. “Policing in Patriarchy: An Experimental Evaluation of Reforms to Improve Police Responsiveness to Women in India.” *Science* 377, no. 6602: 191–198.

Wiswall, M., and B. Zafar. 2015. “Determinants of College Major Choice: Identification Using an Information Experiment.” *Review of Economic Studies* 82, no. 2: 791–824.

Woetzel, J., A. Madgavkar, K. Ellingrud, et al. 2020. *How Advancing Women's Equality Can Add $12 Trillion to Global Growth*. McKinsey & Company.

Wolpe, A. M. 1978. “Education and the Sexual Division of Labour.” In *Feminism and Materialism*. Routledge and Kegan Paul.

Zimmerman, S. D. 2019. “Elite Colleges and Upward Mobility to Top Jobs and Top Incomes.” *American Economic Review* 109, no. 1: 1–47.