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

Ritika Gupta*

September 11, 2023

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 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 50%, 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 a 6% increase in the expected earnings of women.

Keywords: Affirmative Action, Gender, Elite College, IIT, STEM, Subject Choice

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

*PhD Candidate, Dept of Economics, University of Virginia

I am grateful to my advisors Sheetal Sekhri and Amalia Miller for their continual guidance in the project. This paper has also benefited from the participants at the UVA Development Workshop 2021 - Sandip Sukhtankar, Jonathan Colmer, Shan Aman-Rana, Gaurav Chiplunkar and Isaac Mbiti.
This paper was awarded the Best Summer Paper Award 2021 by Dept of Economics at UVA and Huskey Research Grant 2023 by UVA GSAS Council.

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 (Daymont and Andrisani 1984; Beede et al. 2011; Sharpe 1976; Deem 2012; Wolpe 1978; 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 global economy via increased labor market activity and productivity of women, according to a recent research report by McKinsey (Munoz-Boudet and Revenga 2017; Maceira 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 (Bastarrica et al. 2018; Ceci and Williams 2015). 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 a very selective entrance examination which 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 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.

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 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 are twofold - (1) higher

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

wage premium associated with 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 2017-18 from a special round of the National Sample Survey (NSS) focussing 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 kilometer (km) radius of an IIT are considered 'close' whereas districts that lie outside that radius but within 200km are considered 'far'. I use 30km 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

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

reside within 30km of an IIT. 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 (Clarke et al. 2023) 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 science track increased by around 0.02% for every additional seat that was added. This also indicates that the policy has the potential to increase the expected earnings of women by about 6% as the choice of science track is associated with 22% higher earnings on average (Jain et al. 2018), thereby having huge implications in narrowing the gender wage gap.

The results are robust to exclusion of old IITs from the analysis. Dropping one IIT zone at a time from the analysis doesn't affect the results. The results are robust when I restrict sample by redefining far districts to be within 60km, 90km and 120km. I also observe that the effect size and precision reduces as the distance threshold level is increased which corroborates with the conclusion of the conceptual framework. I also conduct a dyadic comparison of close and far districts within each IIT zone and the average coefficient of the same is positive and statistically significant, with magnitude being very close to the triple difference estimate. Finally, the results are also robust to using IIT-zone fixed effects or district by region (rural or urban) fixed effects.

The paper adds to the broad affirmative action literature and attempts to exploit policy variation to study educational outcomes of women, and STEM in particular. Caste-based reservations have been studied in India to determine its targeting and matching properties (Bertrand, Hanna, and Mullainathan 2010; Aygün and Turhan 2017) as well as schooling and educational outcomes of lower-caste students (Bagde, Epple, and Taylor 2016; Khanna 2020). Other studies in the literature look at racial differences with or without using policy variation focusing on affirmative action ban in California and its impact on minority enrolment and attainment (Arcidiacono, Aucejo, and Hotz 2016; Bleemer 2022; Hinrichs 2012; Backes 2012). Finally, studies on affirmative action for women is limited in corporate board leadership and politics (Beaman et al. 2009; Matsa and Miller 2013). The paper distinguishes itself from other related papers that have looked at post college-entry outcomes of the disadvantaged and minority groups by studying choices *women* make *before* entering college in a setting where the college policy provides a natural experiment.

This paper also contributes to the literature on subject choice, which has implications for labor market earnings and the gender wage gap. Jain et al. (2018) establishes that conditional on ability, choosing the science track in high school generates 22% greater earnings for Indian males. This paper investigates whether this choice at high school can be influenced for girls which can potentially reduce the gender wage gap. Moreover earnings associated with elite public colleges are much higher than other colleges (Sekhri 2020; Zimmerman 2019). A wide variety of factors determine choice of subject such as ability, earnings, tastes and preferences (Wiswall and Zafar 2015), role models (Porter and Serra 2020), siblings spillovers (Altmejd et al. 2020) and peer effects (Fischer 2017). Affirmative action literature has studied enrolment, attainment and graduation outcomes but subject choice has been understudied. This paper analyzes if women's choices can be influenced by policies targeted towards them.

This paper also contributes to the literature on 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 gender gap in math and participation in STEM fields (Fryer Jr and Levitt 2010; Adams and Kirchmaier 2016). 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, Niederle, and Oosterbeek

2014; Buser, Peter, and Wolter 2017) and establishing the role of culture in determining math performance (Nollenberger, Rodrıguez-Planas, and Sevilla 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 & 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 are 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 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%

[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

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 2014-15 and 2021-22. 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:

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

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

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

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.

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-squared	0.720	0.725
Number of IITs	20	20

Data Source: Annual Reports of 20 IITs. Gender Ratio is the number of females divided by 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.

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. 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 7,259 girls and 43,781 boys.

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 point in the proportion of women who take the IIT entrance exam as well in the proportion of women clearing the exam, suggesting a 10.5% increase in the proportion of women taking the exam and 15.3% increase in the proportion of women qualifying the exam. In absolute terms, this translates to an yearly average of 4,179 more female registrations and 1,080 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].

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-squared	0.434	0.297
Number of IIT zones	7	7

Data Source: IIT Entrance Exam Reports for years 2013-2020. 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). Independent variable is a dummy taking value 1 for post-policy years.

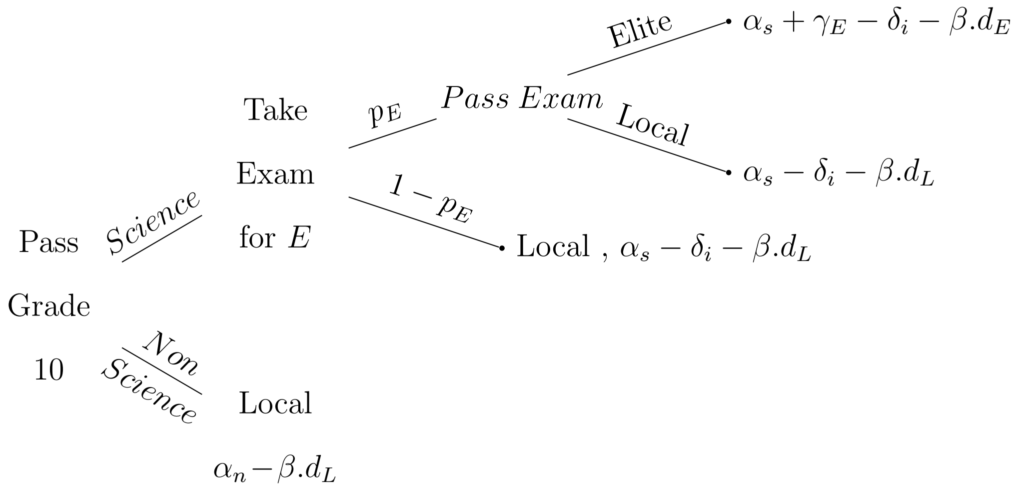
3 Conceptual Framework

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 2019). 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

[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.

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. It also offers all courses. There is, however, the social cost of travelling 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.



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 travelling the extra distance. Mathematically,

$$U_E > U_L$$

$$\implies \gamma_E > \beta(d_E - d_L) \tag{1}$$

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.

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 travelling. 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 science track is greater than the distaste associated with studying science.

$$\implies \delta_i < \alpha_s - \alpha_n \tag{2}$$

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.

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

She chooses science if the extra earnings from 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.

$$\implies \delta_i < (\alpha_s - \alpha_n) + p_E(\gamma_E - \beta(d_E - d_L)) \tag{3}$$

Proposition 1:

a) If a girl lives close (i.e. (1) is satisfied), it is optimal for her to choose science in school as long as her distaste for the subject is not too high (i.e. (3) is satisfied). If she gets into the elite college, she studies a STEM course.

b) If a girl lives far (i.e. (1) is not satisfied), then the choice of subject only depends on the distaste parameter (i.e. choose science if (2) 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 & Identification

4.1 Data

I study the impact of this policy on subject choice by using the 75th round of National Sample Survey (NSS) that focusses on education. The survey was conducted between June 2017 to June 2018 and consists of a nationally representative sample of 64,519 rural households from 8,097 villages and 49,238 urban households from 6,188 blocks. The data covers 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.

Table 3: Descriptive Statistics

	Mean	Std Deviation
All		
Education Level	11.26	1.60
Science	0.29	0.45
Men		
Education Level	11.17	1.52
Science	0.32	0.47
Women		
Education Level	11.38	1.70
Science	0.24	0.43

Notes: The statistics are calculated for the individuals of age greater than 17 years in NSS Education Round 2017-18

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 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 boys to begin with, it is plausible that the trends in the outcome for girls are different than that of boys. I therefore test for the parallel trends assumption for this double difference[7] 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, i.e. 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

[7] The DID estimate corresponding to the double difference is reported in Table 15.

Table 4: Testing the Parallel Trends Assumption

Dependent Variable: Probability of Studying Science		
Panel A: Parallel Trends Assumption for DID		
Age X Female	**-0.0187*** **	
	(0.00330)	
Age	-0.0114***	
	(0.00242)	
Female	0.248***	
	(0.0673)	
Observations	29,105	
R-squared	0.182	
Panel B: Parallel Trends Assumption for DDD		
Age X Female X Close	**0.00127**	**0.000674**
	(0.0101)	(0.0107)
Age X Female	-0.0186***	-0.0189***
	(0.00319)	(0.00406)
Close X Female	-0.0888	-0.0731
	(0.206)	(0.218)
Close X 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-squared	0.182	0.138

Notes: This analysis uses individuals in the 'control' cohort in NSS Education Round 2017-18. 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 parenthesis.

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 choices, gender norms are prevalent and crimes against women are rising[8]. These factors impede female mobility to access schools and colleges and limit their education choices. As presented in Table 5, distance to college matters and it matters more for women and therefore, they travel to colleges closer to their homes. Therefore, girls living closer to IITs have a comparative advantage in accessing such institutions over girls living far

[8] 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.

from IITs. I, therefore, define "close" ("treated") areas as those districts that lie within a 30km radius of an IIT and "far" ("control") areas are the ones that lie outside the 30km radius but within a 200km radius of an IIT[9]. I exclude the districts that are farther than 200km from my analysis to reduce noise and improve precision. Moreover, 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 can not reject the null hypothesis of parallel trends.

Table 5: 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-squared	0.005	0.004

Notes: The data used is the 75th round of National Sample Survey (2017-18) 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 < 1km , 2 for 1km < d < 2kms, 3 for 2kms < d < 3kms, 4 for 3kms < d < 5kms and 5 for d > 5kms.

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. 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. I further

[9] Figure 4 presents a map showing the close ("treated") and far ("control") districts. I choose 30km as my threshold as it is not so high that an average individual cannot commute to college daily or parents cannot 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.

deal with this issue by assigning synthetic difference-in-differences weights (Clarke et al. 2023). This approach is an advanced version of the synthetic control method (Abadie, Diamond, and Hainmueller (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 i.e. 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 synthetic difference-in-difference regression for the main outcome.[10] The method assigns synthetic weights to control districts (those that are farther than 30km) 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 period. 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[11].

5 Estimating Equation & Results

I estimate the triple difference estimate in the following manner:

$$\mathrm{y}_{iaj} = \alpha + \delta\, Female_{ij}\,.\,Young_i\,.\,Close_j + \beta_1\, Female_{ij}\,.\,Young_i + \beta_2\; Close_j.\,Female_{ij} + \beta_3\, Young_i\,.\,Close_j \;+\; \beta_4\, Young_i \;+\; \beta_5\, Female_{ij} \;+\; \beta_6\, Close_j \;+\; \rho\, X_i \;+\; \mu_s \;+\; \epsilon_{iaj}$$

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 kilometers of an IIT & 0 if district j lies within a radius of 200 kilometers of an IIT but farther than 30 kms, μ_j represent the state (or district) fixed effects and X_i denote individual specific controls such as religion, caste, household

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

[11] 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.

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.

Table 6: Triple Difference Analysis

Dependent Variable: Probability of Choosing Science	(1)	(2)
Young X Female X Close	**0.0665*****	**0.0740*****
	(0.0235)	(0.0254)
Female X Young	0.103***	0.0888***
	(0.00767)	(0.0111)
Close X Female	-0.0682***	-0.0802***
	(0.0209)	(0.0238)
Young X 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-squared	0.188	0.159
District FE	Yes	Yes

Notes: Robust Standard Errors clustered at the district level are reported in parenthesis.

I first estimate the equation 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 6 Column 1 provides a triple difference estimate when I choose districts farther than 30km 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. For every additional seat that IIT increased, the likelihood of switching to science track increased by 2.7%.

I also estimate the above equation 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 7. 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.

Table 7: Educational Attainment

Dependent Variable: Educational Level	(1)	(2)
Young X Female X Close	-0.0414	-0.2643
	(0.0813)	(0.2069)
Synthetic DID weights	No	Yes
Observations	59,664	58,592
R-squared	0.529	0.508
District FE	Yes	Yes

Notes: Robust Standard Errors clustered at the district level are reported in parenthesis.

5.1 Robustness & Sensitivity of Results

Excluding the old IITs: Out of the 23 IITs, 7 IITs were established between 1951-1963 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 from my analysis to check if results are driven by these popular IITs. I present the triple difference estimate in Table 8 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 8: Triple Difference Analysis for 16 new IITs

Dependent Variable: Probability of choosing Science	(1)	(2)
Young X Female X 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-squared	0.183	0.156
District FE	Yes	Yes

Notes: Robust Standard Errors clustered at the district level are reported in parenthesis.

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 60km, 90km and 120km 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 30km away with those that are farther. I still observe robust estimates of the policy impact as shown in Table 9.

Table 9: Restricting the Sample

Dependent Variable: Probability of choosing Science	(1)	(2)	(3)
Young X Female X Close	**0.0726** **	**0.0653** **	**0.0634** **
	(0.0299)	(0.0268)	(0.0259)
Observations	14,839	24,029	36,244
R-squared	0.192	0.181	0.179
Band	30km	30km	30km
Far Control	60km	90km	120km
District FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: Robust Standard Errors clustered at the district level are reported in parenthesis.

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 10. I repeat the analysis when a district is close if within 20km, 40km or 50km of an IIT. I find that the impact of the policy is higher when the treated district is within 20km 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).

Table 10: Distance Threshold

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

Notes: Robust Standard Errors clustered at the district level are reported in parenthesis.

Dyadic Comparisons: The main analysis compares districts close to any IIT with those of districts far from any IIT. Additionally, I compare gender gap in science across cohorts between close and far districts *within each IIT zone*. I first divide the districts in the data to 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 gender gap in the likelihood of choosing science before and after the policy between the districts within 30km 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[12] in Table 11. The average coefficient of the individual regressions pertaining to each IIT zone is positive, statistically significant and similar in magnitude with the triple difference coefficient that I obtained in Table 6.

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 12). The use of synthetic DID weights increases the average treatment effect. The results are also robust

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

Table 11: Dyadic Comparison: Average DDD Coefficient

Probability of Choosing Science	(1)	(2)
Young X Female X Close	**0.0672*** **	**0.0606*** **
	(0.0156)	(0.01504)
Synthetic DID weights	No	Yes

Notes: Robust Standard Errors clustered at the district level are reported in parenthesis.

when I include district by rural-urban fixed effects (Table 13).

Table 12: IIT zone Fixed Effect

Dependent Variable: Probability of choosing Science	(1)	(2)
Young X Close X 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-squared	0.117	0.140
Band	30km	30 km
IIT Zone FE	Yes	Yes
Controls	Yes	Yes

Notes: Robust Standard Errors clustered at the district level are reported in parenthesis.

Table 13: Including District by Rural/Urban FE

Dependent Variable: Probability of choosing Science	(1)	(2)
Young X Female X 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-squared	0.197	0.162
District-Rural/Urban FE	Yes	Yes

Notes: Robust Standard Errors clustered at the district level are reported in parenthesis.

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 14. The point estimates vary between 0.053 to 0.076. They vary between 0.058 to 0.089 when SDID weights are used.

Table 14: 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)
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)

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. “Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program”. In: *Journal of the American statistical Association* 105.490 (2010), pp. 493–505.

Adams, Renée B and Tom Kirchmaier. “Women on boards in finance and STEM industries”. In: *American Economic Review* 106.5 (2016), pp. 277–81.

Altmejd, Adam et al. “O brother, where start thou? Sibling spillovers on college and major choice in four countries”. In: (2020).

Arcidiacono, Peter, Esteban M Aucejo, and V Joseph Hotz. “University differences in the graduation of minorities in STEM fields: Evidence from California”. In: *American Economic Review* 106.3 (2016), pp. 525–62.

Aygün, Orhan and Bertan Turhan. “Large-scale affirmative action in school choice: Admissions to IITs in India”. In: *American Economic Review* 107.5 (2017), pp. 210–13.

Backes, Ben. “Do affirmative action bans lower minority college enrollment and attainment?: Evidence from statewide bans”. In: *Journal of Human Resources* 47.2 (2012), pp. 435–455.

Bagde, Surendrakumar, Dennis Epple, and Lowell Taylor. “Does affirmative action work? Caste, gender, college quality, and academic success in India”. In: *American Economic Review* 106.6 (2016), pp. 1495–1521.

Bastarrica, Marıa Cecilia et al. “Affirmative action for attracting women to STEM in Chile”. In: *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*. 2018, pp. 45–48.

Beaman, Lori et al. “Powerful women: does exposure reduce bias?” In: *The Quarterly journal of economics* 124.4 (2009), pp. 1497–1540.

Beede, David N et al. “Women in STEM: A gender gap to innovation”. In: *Economics and Statistics Administration Issue Brief* 04-11 (2011).

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan. "Affirmative action in education: Evidence from engineering college admissions in India". In: *Journal of Public Economics* 94.1-2 (2010), pp. 16–29.

Bleemer, Zachary. "Affirmative action, mismatch, and economic mobility after California's Proposition 209". In: *The Quarterly Journal of Economics* 137.1 (2022), pp. 115–160.

Borker, Girija. "Safety first: Perceived risk of street harassment and educational choices of women". In: *Job Market Paper, Department of Economics, Brown University* (2017), pp. 12–45.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. "Gender, competitiveness, and career choices". In: *The Quarterly Journal of Economics* 129.3 (2014), pp. 1409–1447.

Buser, Thomas, Noemi Peter, and Stefan C Wolter. "Gender, competitiveness, and study choices in high school: Evidence from Switzerland". In: *American economic review* 107.5 (2017), pp. 125–30.

Ceci, Stephen J and Wendy M Williams. "Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men". In: *Frontiers in psychology* 6 (2015), p. 1532.

Chanana, Karuna. "Globalisation, higher education and gender: Changing subject choices of Indian women students". In: *Economic and Political Weekly* (2007), pp. 590–598.

Clarke, Damian et al. "Synthetic difference in differences estimation". In: *arXiv preprint arXiv:2301.11859* (2023).

Daymont, Thomas N and Paul J Andrisani. "Job preferences, college major, and the gender gap in earnings". In: *Journal of Human Resources* (1984), pp. 408–428.

Deem, Rosemary. *Women & schooling*. Routledge, 2012.

Fischer, Stefanie. "The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence". In: *Labour Economics* 46 (2017), pp. 211–226.

Fryer Jr, Roland G and Steven D Levitt. “An empirical analysis of the gender gap in mathematics”. In: *American Economic Journal: Applied Economics* 2.2 (2010), pp. 210–40.

Gupta, Namrata. “Patriarchy reinforced or challenged? A study of engineering students in an elite Indian institute”. In: *Gender, Technology and Development* 24.2 (2020), pp. 250–270.

Hinrichs, Peter. “The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities”. In: *Review of Economics and Statistics* 94.3 (2012), pp. 712–722.

Jain, Tarun et al. “Labor Market Effects of High School Science Majors in a High STEM Economy”. In: (2018).

Khanna, Gaurav. “Does affirmative action incentivize schooling? evidence from India”. In: *Review of Economics and Statistics* 102.2 (2020), pp. 219–233.

Maceira, Helena Morais. “Economic benefits of gender equality in the EU”. In: *Intereconomics* 52.3 (2017), pp. 178–183.

Matheson, Kimberly et al. “Women’s Attitudes Toward Affirmative Action: Putting Actions in Context 1”. In: *Journal of Applied Social Psychology* 24.23 (1994), pp. 2075–2096.

Matsa, David A and Amalia R Miller. “A female style in corporate leadership? Evidence from quotas”. In: *American Economic Journal: Applied Economics* 5.3 (2013), pp. 136–169.

Munoz-Boudet, AM and A Revenga. “Breaking the STEM ceiling for Girls”. In: *Brookings. Retrieved on January* 14 (2017), p. 2018.

Niederle, Muriel and Lise Vesterlund. “Explaining the gender gap in math test scores: The role of competition”. In: *Journal of Economic Perspectives* 24.2 (2010), pp. 129–44.

Nollenberger, Natalia, Núria Rodrıguez-Planas, and Almudena Sevilla. “The math gender gap: The role of culture”. In: *American Economic Review* 106.5 (2016), pp. 257–61.

Porter, Catherine and Danila Serra. “Gender differences in the choice of major: The importance of female role models”. In: *American Economic Journal: Applied Economics* 12.3 (2020), pp. 226–54.

Resmini, Marina. *The ‘leaky pipeline*. 2016.

Sekhri, Sheetal. “Prestige matters: Wage premium and value addition in elite colleges”. In: *American Economic Journal: Applied Economics* 12.3 (2020), pp. 207–25.

Sharpe, Sue. *’JUST LIKE A GIRL’.: How Girls Learn to be Women*. Penguin Books, 1976.

Wiswall, Matthew and Basit Zafar. “Determinants of college major choice: Identification using an information experiment”. In: *The Review of Economic Studies* 82.2 (2015), pp. 791–824.

Woetzel, J et al. *How advancing women’s equality can add $12 trillion to global growth. McKinsey & Company*. 2020.

Wolpe, Ann Marie. “Education and the sexual division of labour”. In: *Feminism and materialism. London: Routledge and Kegan Paul* (1978).

Zimmerman, Seth D. “Elite colleges and upward mobility to top jobs and top incomes”. In: *American Economic Review* 109.1 (2019), pp. 1–47.

6 Appendix

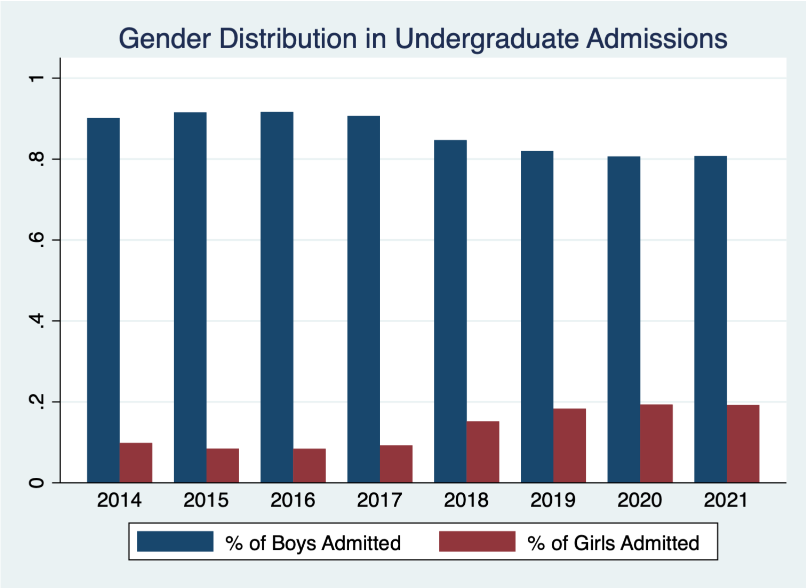
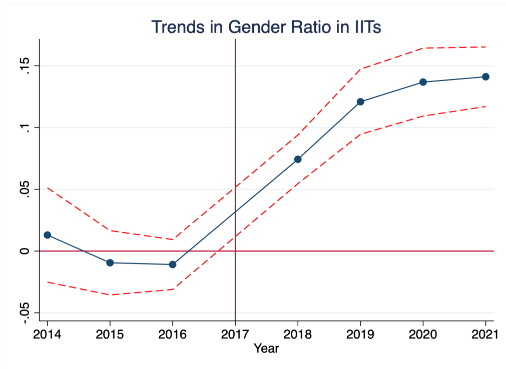


Figure 1



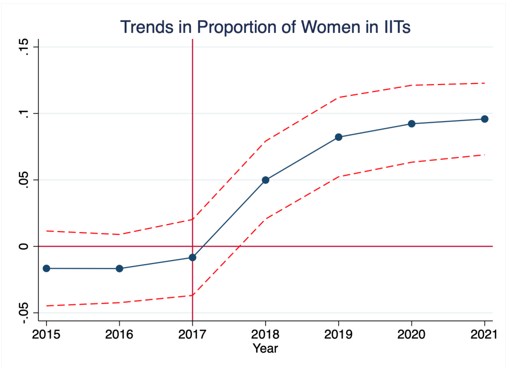


Figure 2: Trends based on numbers from 20 IITs between 2014-2021

Figure 3: 23 IIT Campuses

Table 15: Difference-in-differences estimate

Dependent Variable: Probability of Studying Science	(1)
Young X Female	**0.113*** **
	(0.00737)
Young	-0.188***
	(0.00896)
Female	-0.127***
	(0.00645)
Observations	59,664
Control Mean	0.24
R-squared	0.187
District FE	Yes
Controls	Yes

Notes: Robust Standard Errors clustered at the district level are reported in parenthesis.

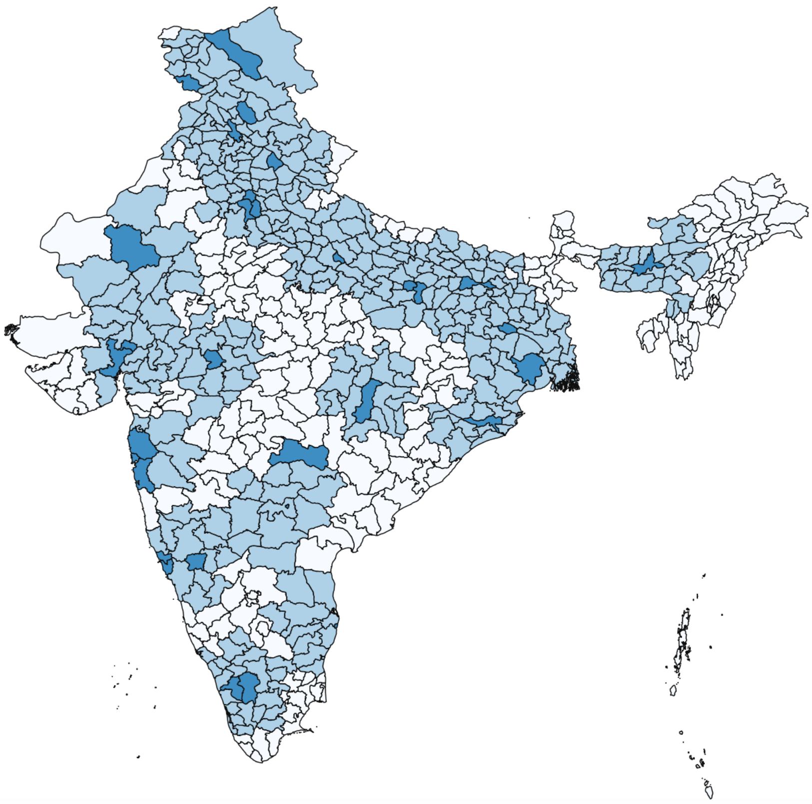


Figure 4: Close: Districts that come under 30km radius of an IIT Campus
Far: Districts that come under 200km radius (but more than 30km) of an IIT Campus