APRIL 14, 2022

WHAT CAN(NOT) EXPLAIN THE GAP?

Evidence and Decomposition of Gendered Stream Choice in India

WORKING PAPER

TRIDIP RAY; ARKA ROY CHAUDHURI AND KOMAL SAHAI

O P JINDAL GLOBAL UNIVERSITY

https://jgu.edu.in/

WHAT CAN(NOT) EXPLAIN THE GAP?

Evidence and Decomposition of Gendered Stream Choice in India

TRIDIP RAY[†] ARKA ROY CHAUDHURI[‡] KOMAL SAHAI[±]

April 14, 2022

Abstract

Gendered pattern in stream choices is well established in the education literature. Males are over represented in the mathematically oriented courses while females are more likely to opt for life sciences and non-science courses. We use three cohorts of student results data from the Central Board of Secondary Education, the single largest education board with an all-India presence, to first quantify and subsequently decompose the gender gap in the very first stream choices made by students at the school level in India. We use our rich dataset to explore a large set of explanatory factors proposed in the literature to explain the gender gap, namely, student ability, attributes of cohort peers, "chilly" climate and socioeconomic characteristics of students. We employ a novel way to use a student's immediate seniors in schools to elicit the "chilly" climate aspect of stream choice. Our measure of the expected "chilliness" of the climate in a prospective course accounts for the largest portion of the observed gender gap in male dominated subjects. It explains 13.48% of the gap in the take-up of Mathematics. Ability related attributes explain none of the gender gap in Mathematics, but 12% of that in Biology. The contribution of peer related variables, on the other hand, is negligible.

KEYWORDS: Stream choice; gender gap; chilly climate; decomposition; India

JEL codes: I20, I21, J16, J24

[†]Professor, Economics and Planning Unit, Indian Statistical Institute, New Delhi; tridip@isid.ac.in
[‡]Assistant Professor, Department of Economics, Shiv Nadar Univeristy, Delhi, NCR; arka.chaudhuri@snu.edu.in
[±]Assistant Professor, Jindal School of Gvernment and Policy, Jindal Global University, Sonepat; ksahai@jgu.edu.in

1 Introduction

Gender gap in earnings is well established in the economics literature, both in the context of developed[1] as well as developing countries[2], including India[3]. Even in 2014, annual earnings of full-time female workers in the USA were only about 79% that of male workers (Blau and Kahn 2017), while the cross-country unadjusted gender gap in average wages in Europe was 14.2% (Boll and Lagemann 2018). The gap stood much higher at 49% in India in the year 2009-10 (Deshpande et al. 2018). While a number of explanations have been extended to explain this gap, occupational segregation has emerged as a major explanatory factor[4]. In particular, male dominated occupations related to Science, Technology, Engineering and Mathematics (STEM) fields have substantial earnings premium while women are over represented in lower paying jobs like nursing and teaching[5]. Why are so few women employed in STEM related occupations despite a clear economic advantage in these fields? It could be either because fewer women graduate in STEM or because women graduate in STEM but drop out of STEM occupations, or a combination of both. In most countries, however, specialization into STEM, or more broadly, into Science and non-Science streams happens even earlier, at the school level itself. In this paper, we look at the very first stream choices made by students at the school level in the context of India

[1]See, for example, Rubery (1992); O'Neill (2003); Plantenga et al. (2006); Blau and Kahn (2017); Boll and Lagemann (2018).

[2]See, for example, Ashraf and Ashraf (1993); Brown et al. (1999); Maurer-Fazio et al. (1999); Rendall (2013); Chi and Li (2014); Guimarães and Silva (2016).

[3]See, for example, Reilly et al. (2005); Menon and Van der Meulen Rodgers (2009); Khanna (2012); Das (2012); Duraisamy and Duraisamy (2016); Deshpande et al. (2018).

[4]See, for example, Dolado et al. (2002); Hegewisch et al. (2010); Hegewisch and Hartmann (2014); Blau and Kahn (2017).

[5]See James et al. (1989); Grogger and Eide (1995); Dolton and Vignoles (2002); McGuinness (2003); Buonanno and Pozzoli (2009); Webber (2016); Belfield et al. (2018); Dahl et al. (2020).

using three cohorts of students under the single-largest education board[6] with an all-India presence – the Central Board of Secondary Education (henceforth CBSE).

A rich literature is devoted to exploring the college major choice of students[7]. Recent work has shifted focus to look at stream choices made earlier, at the school level[8]. In the Indian context, however, the literature is patchy and limited in scope. While Mattoo (2013) and Gautam (2015) have limited geographic scope, Chanana (2000, 2007) only look at aggregate statistics. Chakrabarti (2009) and more recently, Prakasam et al. (2019) use nationally representative data sets to look at major choices at the college level. In a recent paper, Sahoo and Klasen (2020) look at stream choices at the school level using the nationally representative Indian Human Development Survey. To the best of our knowledge, ours is the first paper to use multiple cohorts of a large scale dataset to establish and subsequently decompose the gender gap in stream choice in India. Using regression and linear decomposition techniques, we decompose this gap to estimate how much of it can be "explained" by an expansive set of explanatory factors that have been proposed in the economics, sociology and psychology literature. In particular, we evaluate how much of the gender gap is accounted for by a difference in student ability, their cohort peers in school, their immediate seniors in school and their socioeconomic status. Our results highlight two nuances: One, there is a divide within the STEM subjects and the gender gap in the take-up exists along the lines of life sciences and physical sciences. Two, the relative importance of each of the explanatory factors depends on the dominating

[6]An education board in India is defined by its jurisdiction. There are three national boards and several state boards in India (Cheney et al. 2005).

[7]See, for example, Turner and Bowen (1999); Montmarquette et al. (2002); England and Li (2006); Dickson (2010); Riegle-Crumb and King (2010); Riegle-Crumb et al. (2012).

[8]See, for example, Baram-Tsabari and Yarden (2011); Ajayi and Buessing (2015); Friedman-Sokuler and Justman (2016); Justman and Méndez (2018); Rapoport and Thibout (2018); Landaud et al. (2020).

gender in a given subject. Ability related variables are the largest contributors to the explained part for the female dominated subject Biology. On the other hand, for the male dominated subjects like Mathematics, our measure of chilly climate emerges as the most important factor discouraging girls from opting for the subject.

Since we are looking at choices made at the school level, the classification of subjects into STEM and non-STEM can be misleading for two reasons. First, the subjects offered after matriculation are more basic like Physics and Chemistry rather than Engineering. Second, "Science" in STEM incorporates both Technical sciences (like Engineering) and Life sciences (like Medicine and Anthropology), but the distinction is not sharp at the school level, again because the subjects offered are at the basic level. For example, almost 99% of students in our dataset who opt for Biology opt for Physics and Chemistry as well. Therefore, we find it more prudent to look at individual subjects and subject combinations most commonly offered by the schools. In particular, we look at three subjects/subject combinations in this paper: Mathematics, Physics-Chemistry-Mathematics (PCM) combination, and Biology[9]. Our choice is driven by the fact that not only are the gender differences starkest in these subjects, but they also bring out the physical sciences-life sciences divide within the Science stream.

We begin by quantifying the gender gap in various subjects offered in schools after matriculation and find a clear gender divide in our dataset along the same lines as observed in the literature. Boys are 19.13 percentage points more likely than girls to take-up Mathematics in class XI and 20.61 percentage points more likely to take-up PCM. Girls, on the other hand, are 11.18 percentage points more likely than boys to take-up Biology. In general, a higher proportion of girls takes

[9] It must be noted that Mathematics can be a subject choice even without PCM; for example, it can be chosen with the Commerce stream (combination of Business Studies and Accounts) as well.

up Biology, Economics, Political Science and History, while boys are more likely to take-up PCM and Computer Science. It is important to highlight that even though a higher proportion of girls takes up Biology, the overall take-up of “Science” subjects among girls is still lower than boys.

Next we proceed to dissect this gap. Facilitated by our rich dataset, we go beyond student ability and explore the literature for other plausible reasons extended for the gender gap in stream choice. We group them into four broad heads: Ability, Cohort peers at school, Immediate seniors at school and Socioeconomic characteristics. For each of these heads, we first show descriptive statistics on how it is distributed between boys and girls. Then we examine them under a regression framework using Linear Probability Models and a decomposition framework using Oaxaca Blinder decomposition technique.

Based on some early life studies which find gender gap in Mathematics scores favoring boys[10], this difference is often cited as the explanation for the observed gender divide in stream choice. This claim has been falsified in multiple studies in the context of developed countries (Dickson 2010; Riegle-Crumb and King 2010; Riegle-Crumb et al. 2012; Rapoport and Thibout 2018; Friedman-Sokuler and Justman 2016; Justman and Méndez 2018). We test this in the context of a developing country using a student’s class X score as a proxy for her ability. A more nuanced explanation explored in the literature is comparative advantage in the relevant subject. It is argued, for example, that boys have a comparative advantage in Mathematics over languages while girls have a comparative advantage in life sciences over other technical sciences, and thus we observe the existing gender divide in stream choices (Park et al. 2007; Valla and Ceci 2014). For this, we use class X scores in the rele-

[10]See, for example, Penner and Paret 2008; Fryer Jr and Levitt 2009; Wai et al. 2010.

vant subjects to construct comparative advantage variables. These two variables are considered under the ability head of our explanatory variables.

Under the cohort peers head, we explore the argument that males and females may have very different responses to a given set of peers (Gneezy et al. 2003; Gneezy and Rustichini 2004; Niederle and Vesterlund 2007; Gneezy et al. 2009; Fletschner et al. 2010). The reasons for this are said to be rooted in differential confidence[11] and risk taking behaviour [12] across genders. To see this, we include cohort composition variables and cohort performance variables at the school-cohort level under this category.

Next, we employ a novel way to utilize the immediate seniors that students had in class X to elicit another possible explanation proposed in the sociological and psychological literature: the phenomenon of "Chilly Climate"(Clark Blickenstaff 2005). It says that females are shy of choosing male dominated fields because they expect to face a hostile environment there. If too few fellow females are present in a Mathematics class, then there are higher chances of covert and overt discrimination, or a general feeling of being at a loss[13]. The same could be applicable to males in a Biology class dominated by females. We propose that students may form an idea of a prospective Mathematics class, for example, by looking at the gender composition of the Mathematics class of their immediate seniors in school. To the best of our knowledge, ours is the first paper to use information on a student's seniors to study the chilly climate aspect in stream choice.

Finally, we also probe how much of the gender gap can be attributed to socioeconomic characteristics of students. We use their caste status[14], annual family income

[11]See Jakobsson (2012); Pirinsky (2013); Sarsons and Xu (2015).

[12]See Charness and Gneezy (2012); Hardies et al. (2013).

[13]See, for example, Sadker and Sadker (1986); Fouad et al. (2011); Lordan and Pischke (2016); Tellhed et al. (2017); Wu (2017).

[14]Caste is a system of social stratification in the Indian society based on the ancient *varna* system

and single child status as variables signaling their socioeconomic status.

Our descriptive statistics show that the distribution of absolute ability does not differ dramatically across girls and boys, nor do the attributes of cohort mates at school. In contrast, there are large differences between the genders when it comes to the measure of chilly climate. Girls expect a much more chillier climate in Mathematics and PCM compared to boys. The opposite is observed for girls and boys in Biology.

Next, we test these explanatory factors under the Blinder-Oaxaca decomposition framework. We first include the four broad heads one by one, then test their strength against each other by including them together. We report three broad findings. One, our measure of chilly climate is the largest explanator of the gender gap in the male dominated subjects of Mathematics and PCM. If girls had the share of own gender students in the senior Mathematics and PCM classes like boys, thus expecting a "warmer" climate like boys, the gender gap in these subjects would have closed by 13.48% and 14.73%, respectively. Two, for Mathematics and PCM, a comparative advantage in Mathematics vs English is more important in explaining the gender gap than any measure of absolute ability. For Biology, on the other hand, class X total score is the most important factor. If girls had the (lower) average scores in class X as boys do, they would be 12.08% less likely to opt for Biology and thus the gap would be smaller. Three, cohort composition and cohort performance variables, do

(Deshpande 2011). The castes used in this paper are administrative categories that are used for all official work in India. They are the Scheduled Castes (SC) or the former "untouchables" who are situated at the bottom of the social hierarchy; the Scheduled Tribes (ST), who are the backward and marginalised tribes of India; the Other Backward Classes (OBC), which comprise of other backward, but not "untouchable" castes, and General or the residual category, comprising of castes at the top of the hierarchy. Though originally a Hinduism construct, all religions in India are known to follow some sort of caste system in their communities (Ray et al. 2020). See Deshpande (2011) for a detailed discussion of the caste system, its evolution and its socio-political relevance in present times.

not explain any statistically significant portion of the gender gap in any of the three subjects.

Our contribution to the literature is fourfold. First, we use a newly available administrative results dataset of the census of students under the largest national level education board in India. The rich student level data allows us to study fine differences within a broad stream. This is particularly important in the Indian context where most secondary large datasets have information only at the broad stream level like Science, Arts and Commerce. Similarly, where most datasets only have information on the overall grades (first or second division) and results (pass/fail) of students, we know the exact subject-wise scores of each student. This gives an upper edge in exploring individual level outcomes using individual level controls.

Second, while substantial work has been done on stream choice in developed countries, most papers focus on only a particular explanatory factor, like ability or classroom peers. Our paper is the first one to present multiple factors covering a broad spectrum of explanations in a unified framework empirically. We utilize an expansive set of observable factors around a student to account for the gender gap present in stream choice.

Third, we implement a novel way to elicit the chilly climate aspect of stream choice using school seniors of students. Previous work has only used the composition of classroom peers to measure "chilliness" of the chosen course of study. To the best of our knowledge, ours is the first paper to use a student's seniors, whom the students observe *before* they make their own decisions regarding stream choices, to construct this measure.

Fourth, ours is also the first paper to rigorously establish and subsequently decompose the gender gap in stream choice at the school level using multiple cohorts of student level data in the context of India. The existing work in the Indian context is patchy and limited in geographic scope. We are able to fill this gap using rich detailed data of over 2 million students.

The rest of the paper is organized as follows. In Section 2 we describe the institutional background followed by a description of the dataset in Section 3. In Section 4 we provide a descriptive analysis with respect to each explanatory factor and each of the three subjects. Section 5 briefly discusses the methodologies used in the paper. Section 6 shows the results of the regression and the decomposition exercises. Section 7 offers a brief discussion of the results and its implications, and Section 8 concludes.

2 Institutional Background

The education system in India follows a 10+2+3 structure recommended by the National Policy on Education, 1968. It comprises of 10 years of schooling up to the secondary level culminating into matriculation in class X. This is followed by 2 years of higher secondary schooling in class XI and XII, and then 3 years of graduation. In general, the duration of graduation varies depending on the type of course and degree. The duration of schooling, however, uniformly follows the 10+2 pattern across the country (Cheney et al. 2005).

A school in India is affiliated to a board of education. A board is defined by its jurisdiction and follows a common curriculum across all affiliated schools[15]. There are three national and several state boards in India (Cheney et al. 2005). The Central Board of Secondary Education (CBSE) is the single-largest education board in the

[15] Though a school's syllabus is the responsibility of the board it is affiliated to, in theory, it must be aligned with the National Curriculum Framework, 2005 (Anderson and Lightfoot 2019).

country with an all-India presence. As of 2019, there were 21271 schools affiliated to CBSE in India and 228 schools in 25 foreign countries (cbse.nic.in).

All schools under the CBSE follow the same curriculum till class X, except for the choice of languages (http://cbseacademic.nic.in/curriculum_2021.html)). All education boards conduct board level standardized examinations at the end of secondary school in class X, and then at the end of higher secondary school in class XII. These examinations, commonly called board examinations, have common question papers and evaluation guidelines across the board. Both the board examinations, especially the class XII board examinations, are high stake examinations because their results are used for admission to various colleges and institutions for higher education.

Under the CBSE, after studying a common syllabus till class X (two languages, Mathematics, Science[16] and Social Science (http://cbseacademic.nic.in/curriculum_2021.html)), students have to choose one language and four specialized subjects for the next two years of study (classes XI and XII). CBSE offers a wide range of scholastic and co-scholastic subjects to choose from[17]. The choices offered by a school, however, may be limited due to resource constraints on part of the school. The most common subjects opted by students in our dataset are Mathematics, Physics, Chemistry, Biology, Computer Science, History, Political Science, Geography, Economics, Hindi, English, Business Studies and Accounts. Though a student can choose any set of subjects from the available options, there are some combinations chosen most frequently. These include Physics-Chemistry-Mathematics (PCM), Physics-Chemistry-Biology (PCB), History-Political Science (Arts) and Business Studies-Accounts (Commerce).

[16] It may be noted that till class X, students are taught the subject "Science". This is then split up into Physics, Chemistry and Biology in class XI. Throughout the paper we use Science to mean the subject taught in class X.

[17] For the complete list of subjects offered by CBSE, please refer to http://cbseacademic.nic.in/curriculum_2021.html.

3 Data

We use three cohorts of newly available results data from the Central Board of Secondary Education in India. The board conducts two national level standardized examinations every year, one for class X and the other for class XII students. Over 1.2 million students appear for class X board examinations annually, while over 1 million students appear for class XII board examinations on an average. The stream of study chosen after matriculation is a major determinant of the field of study a student can choose in college or university taking admission via various national and state level competitive examinations. We study the stream choices of three cohorts of students who took the standardized board examinations under the CBSE at both class X and XII level.

For each student, we have board examination results data for each subject in class X (2 languages, Mathematics, Science and Social Science) and for each subject in class XII which they opted for after class X. We have their exact scores out of 100 in each of these subjects as well as their grades ranging from A to E, along with the overall score, grade and final result (pass/fail). Thus, for each class XII student, we know the subjects they opted for their higher secondary schooling and their scores in those subjects. We match these class XII students to their class X results using a unique roll number assigned to each class X student in each academic year. From a total of 3,134,622 class XII students in three cohorts from years 2014, 2015 and 2016, 2,406,163 are matched with their class X results from years 2012, 2013 and 2014, respectively[18]. We restrict our analysis to this set of students who appeared for both class X and XII board examinations under the CBSE.

[18]The unmatched 728,459 or 23.24% students, that is, the students who did not have a CBSE class X roll number, are the students who have migrated to CBSE in class XI from other school boards.

In addition to subject scores and grades, we also have information on the gender, date of birth, caste status, annual family income and single child status of students. The data also tells us the type of their school administration (public, private, and so on). We also have a school identifier for each student at both class X and XII, which need not necessarily be the same.

Table 1 reports the summary statistics for the variables available in the CBSE data. Out of a total of 2,405,349 students, around 44% are girls. The mean age of students in class X is 16.68 years. Almost 73% of the students belong to the unreserved General caste category, while 7.31% belong to Scheduled castes (SC), 3.32% to Scheduled tribes (ST) and the remaining 16.31% belong to Other Backward Classes (OBC).

In class XII, the largest majority of students (68%) are enrolled in private schools[19]. The Government schools come second enrolling close to 20% of students in our sample. More than 62% of class X students opt for school based board examination. The mean annual family income of students is 292,655.8 INR with a very large variance. Around 6% students are single children in their families. The mean score of students in class X is 348.91 out of 500 and in class XII is 329.93 out of 500.

4 Descriptive Analysis

In this section, we discuss in detail each of the explanatory factors we consider for the decomposition exercise. We describe the variables used to measure the factors

[19]Schools in India can be government owned (Government), privately owned but financially aided by the government (Private Aided), or privately owned by individuals, trustees or societies (Independent)(Anderson and Lightfoot 2019). The Kendriya Vidyalayas are government owned special schools to cater to the educational needs of the children of transferable Central Government employees. The Jawahar Navodaya Vidyalayas are also government owned schools for talented students predominantly from rural areas.

and their differential distribution across girls and boys.

4.1 Overall stream choice

Though we take an in-depth look into the choice of only three subjects in this paper – Mathematics, Biology and the Physics-Chemistry-Mathematics (PCM) combination, this subsection gives a general view of the gender divide with respect to all subjects in class XI.

Table 2 reports the pattern of take-up of the most common subjects offered to CBSE students after matriculation. Column 1 specifies the subject/subject combination, column 2 shows the percentage of all students who opted for that subject in class XI. column 3 shows what proportion of those who opted for a subject were girls. Finally, columns 4 and 5 show the shares of girls and boys, respectively, who opted for the given subject.

The gendered pattern in the choice of subjects comes out clearly from Table 2. While more than 53% of boys take-up Mathematics after class X, only around 35% of girls do so. The gap is slightly muted but still close to 13 percentage points in case of Physics and Chemistry, and around 5.5% in Computer Science. In contrast, the gender gap favours the girls in case of Biology, History, Political Science and Hindi. It is the largest, at around 12 percentage points, for Biology and around 7 to 8 percentage points for History, Political Science and Hindi. For the rest of the subjects, namely Geography, Economics, English, Business Studies and Accounts, the difference in the take-up rate between girls and boys is 5 percentage points or below.

The last three rows of Table 2 show subject combinations[20]. It can be seen that

[20]It must be reiterated that these subject combinations have been created by the authors to match the most commonly taken subject combinations in CBSE. The board does not require students to

there is an over 20 percentage points gap in the take-up of PCM in favour of boys. The difference is much smaller, and in favour of girls, in case of Arts and Commerce: 5 percentage points and 2 percentage points, respectively.

It is clear from Table 2 that the largest differences in take-up between girls and boys exists in Mathematics and PCM in favour of boys, and in Biology in favour of girls. From here onward in this paper, we focus on these three streams only. We now move on to discuss each of the four categories of explanatory factors in detail.

4.2 Ability

It was briefly discussed in the introduction that multiple studies have established that a difference in ability across girls and boys now accounts for a negligible portion of the gender gap in stream choices (Turner and Bowen 1999; Dickson 2010; Riegle-Crumb and King 2010; Riegle-Crumb et al. 2012; Friedman-Sokuler and Justman 2016; Rapoport and Thibout 2018; Justman and Méndez 2018). As a matter of fact, as we progress towards more recent studies, lesser and lesser portion of the gender gap is explained by a difference in ability. For example, Turner and Bowen (1999) find, using a Blinder Oaxaca decomposition, that Scholastic Aptitude Test (SAT) scores explain as much as 45% of the gender gap in the take-up of Mathematics-Physical Sciences as college major in the USA. In contrast, using the same decomposition technique, Justman and Méndez (2018) show that Australia's National Assessment Program-Literacy and Numeracy (NAPLAN), along with other control variables does not explain the gap in Physics, Information Technology and Specialist Mathematics at all.

Table 3 shows score comparison between girls and boys in our dataset. We use

opt for subjects from a combination package and they can choose any five subjects from the available subjects.

each student's class X score in the CBSE board examination as a proxy for their ability. Since students make subject choices immediately after class X, these scores make justifiable candidates. In addition, they are standard across all schools under CBSE. Thus they can be used even if students changed schools post matriculation. There are five subjects in class X: Mathematics, Science, Social Science, and two languages. While 99.83% of students have English as one language in class X, 83.76% of students have Hindi. Table 3 reports the scores in these five subjects for girls and boys. Column 2 gives the overall means, column 3 gives the means for girls and column 4 for boys. The last column gives the difference between the means for girls and boys in terms of the standard deviation of the overall sample (value in column 2).

It is evident from Table 3 that the differences between the average scores of girls and boys are very small: only a small fraction of the corresponding standard deviations. In fact, girls outperform boys in every subject, except in Mathematics, where they lag behind by 0.04 of a standard deviation.

Next, Figure 1 explores how the class X scores relate to students' stream choices. For each graph in Figure 1, we calculate score deciles for boys and girls separately. Then we calculate the proportion of boys (girls) in each decile who opted for the given subject in class XI. These proportions are then plotted against the class X score deciles.

In the top panel, Figures 1a, 1b and 1c show the rate of take-up of Mathematics, PCM and Biology over class X *total* score deciles. Firstly, the take-up increases with score for both girls and boys for all three streams, implying that ability is a positive predictor of these subjects in general. Secondly, the gap in stream choice exists at virtually all deciles of the total score distribution, strongly implying that total scores will be poor explanators of the gender gap in stream choice. For Mathematics and PCM, it starts low at the lower deciles, then stabilizes around the 4^{th} decile. For Biology, the gap is flipped in favour of girls and is practically zero till the 2^{nd} decile. Beyond that the gap increases with increasing score, mainly because the curve for boys flattens after around the 6^{th} decile.

In the next two panels, we replace the class X total score deciles with score deciles of the relevant subject in class X to see if these do a better job of explaining the gender gap. Figures 1d and 1e in the second panel plot the gender wise take-up rates of Mathematics and PCM over class X *Mathematics* score deciles. We see a very similar pattern as seen in the previous panel: the gap starts low, stabilizes around the 4^{th} decile, and exists throughout the Mathematics score distribution. The last panel shows the take-up rates for PCM and Biology over class X *Science* score deciles. Here too, the graphs resemble the ones in the top panel. Figure 1g for Biology, for example, shows that the gap is negligible in the beginning and widens with increase in score.

The discussion so far in this section suggests that while ability variables are positively correlated with take-up of Mathematics, PCM and Biology, they are unlikely to account for a major portion of the gender gap. This is because the score distributions do not differ substantially between genders and moreover, the gender gap exists at every point in the score distribution. A more nuanced variant of the ability related explanations is the concept of Comparative Advantage in ability which we take up next.

Comparative advantage in ability borrows from the concept of comparative advantage in international trade. It means that boys do better in Mathematics and mathematically oriented subjects *compared* to subjects involving language and verbal skills. Girls, on the other hand, have an advantage in language and life sciences. It is this comparative advantage in ability that drives the gendered pattern of stream choices observed across countries (Park et al. 2007; Valla and Ceci 2014). Many recent studies have looked at the role such differential ability advantage plays in students' stream choices and found that the differences are more cultural than biological. For example, Friedman-Sokuler and Justman (2016) find no evidence that boys' comparative advantage in Mathematics drives their higher take-up of the subject in a sample of Israeli schools. Justman and Méndez (2018) show that comparative advantage exists but not at the gender level, rather between students with and without an English background in Victoria, Australia. In fact, Pope and Sydnor (2010) show that comparative advantage varies significantly across the US, suggesting again that culture and environment greatly affect the observed differences in abilities. In a recent paper, Goulas et al. (2020) exploit the random assignment of students to classrooms in the context of Greece and find that for females, comparative advantage matters when it is higher as compared to their classroom peers, while it seems to be irrelevant for boys.

The source of the difference notwithstanding, existence of comparative advantage in a subject over others may still drive stream choice across genders. Table 4 quantifies the comparative advantage in ability in our data. Since we are comparing across subjects, we convert absolute scores into standardized scores[21]. The first set of rows in Table 4 gives the mean standardized scores of class X scores for girls and boys. As also seen in Table 3, girls score above the mean in all subjects except in Mathematics. For the mean gap in scores in the next set of rows, we simply subtract the standard-

[21]Scores are standardized by subtracting the cohort mean from a student's own score and dividing the difference by the cohort standard deviation. The resultant standardized scores will have a mean of 0 and standard deviation of 1 within a cohort.

ized score in English, for example, from that in Mathematics and average it over girls and boys separately. We see that boys indeed have a statistically significantly higher comparative advantage in Mathematics and Science over language compared to girls: boys score better in Mathematics and Science compared to English on an average (difference is positive). The opposite is true for girls. It is interesting to note that boys also have a comparative advantage in Mathematics compared to Science. Girls do better in languages compared to Mathematics and Science, but they do better in Science compared to Mathematics as well. This could potentially account for some of the gap in the take-up of Biology, the non-mathematical Science course. Finally, the last set of rows show that the above pattern exists in the overall sample as well: a higher proportion of boys secured higher scores in Mathematics and Science than in English, while a higher proportion of girls secured higher scores in Science than in Mathematics.

To summarise, while ability correlates positively with a higher take-up of Mathematics, Biology and PCM, the gender gap in the take-up is undiminished across the entire distribution of scores. On the other hand, we see that boys do have a comparative advantage in Mathematics over languages as well as Science.

4.3 Cohort Peers

A number of studies have shown that behavioural differences exist between males and females. Females have been shown to exhibit less confidence, lower competitiveness and higher risk aversion (Gneezy et al. 2003; Gneezy and Rustichini 2004; Niederle and Vesterlund 2007; Gneezy et al. 2009; Fletschner et al. 2010; Charness and Gneezy 2012; Jakobsson 2012; Hardies et al. 2013; Pirinsky 2013; Sarsons and Xu 2015). An implication of these differences is that men and women may behave differently among the same peers. Undoubtedly, the education literature has built upon this observation and explored whether peer composition affects female performance differentially (Marsh et al. 2008; Hunt 2016; Anelli and Peri 2019; Fischer 2017; Kugler et al. 2017; Bostwick and Weinberg 2018; Astorne-Figari and Speer 2019; Landaud et al. 2020). Hunt (2016) and Bostwick and Weinberg (2018) show, for example, that a lower proportion of females among peers can negatively impact a woman's outcomes. Fischer (2017) and Landaud et al. (2020), on the other hand, show that higher performing peers can be detrimental to female performance and choices. We explore these two aspects of peers in our context.

First, Table 5 describes the gender composition in our data. We look at gender composition at class X level since stream choice decisions are potentially based on factors present at that time. Panel A reports that out of a total of 12,691 schools in our data, 1,046 are all-girls schools, 1,347 are all-boys schools and a majority 11,253 are co-educational schools. The average female share in cohort is 44% in the overall sample and close to 42% in the sample of co-educational schools. Panel B shows the gender composition of peers in the cohort of an average girl and boy in the data. Overall, girls have a higher proportion of girls in their peer group at school, and boys have higher proportion of boys. This is to be expected given the existence of single-sex schools in the data. Thus we look at only co-educational schools in the last two rows. Here, while the average female share is below 50% for both girls and boys, the average girl has 46% female share in her cohort, while the average boy has only 39% females in his cohort.

We now consider the second aspect of cohort peers: peer performance[22]. Here

[22]Peer performance is calculated by subtracting a student's own score from the sum total of cohort score of the school and then dividing the difference by the number of peers (total number of students in the school minus 1). Gender wise peer performance indicators are calculated similarly, except that the student's own score is subtracted only when own gender matches the gender of the

too we look at class X cohort peers of students and their performance in class X board examination. Panel A of Table 6 reports mean peer performance figures for all schools in the data. The peer set of girls are lower performing than that of boys. An average peer of a girl scores 345 out of 500 in class X, while that of a boy scores 352 on average. The next two rows depict performance figures for female and male peers separately. Again, boys have higher performing female peers than girls but slightly lower performing male peers than girls.

However, looking at only male and female cohort peers at a time eliminates single-sex schools for at least one gender[23]. Thus, we report the numbers for co-educational schools only in Panel B. Here we see that girls have an overall lower performing set of both female and male peers.

4.4 Immediate Seniors

In this section, we build on the premise that a student interacts with a number of people in a school environment. This includes her teachers, cohort peers as well as school seniors. Under this subsection, we use a student's immediate seniors when in class X to elicit an explanatory factor called "Chilly Climate" (Clark Blickenstaff 2005), which is novel in the economics literature.

We have three consecutive cohorts in our dataset: class XII batches of 2014, 2015 and 2016 (who were in class X in 2012, 2013 and 2014, respectively). The class X of 2012 (2014 batch of class XII) is the immediate senior of the class of 2013, and the class X of 2013 is the immediate senior of the class of 2014. As we do not have data on class X batch of 2011, we only consider the later two cohorts for this part of the

peer group being considered.

[23]For example, looking at only female cohort peers eliminates only-boys schools and vice versa when looking at only male cohort peers.

analysis, and use the two older cohorts for constructing our independent variables involving immediate seniors. Since we are looking at decisions made after class X, we consider seniors in the schools students were enrolled in class X. A number of studies have documented that when women enter male dominated fields of study or occupation, they encounter a hostile or an unwelcome environment or a "chilly climate". A male dominated field translates into overt or covert discrimination, a feeling of misfit or being at a loss (Sadker and Sadker 1986; Fouad et al. 2011; Lordan and Pischke 2016; Tellhed et al. 2017; Wu 2017). Potentially, this is also a possibility for males entering female dominated fields. We propose that students can gather an idea about what it would be like to enter into streams dominated by the other gender by looking at their seniors. In particular, students in class X can look at the gender composition of the senior Mathematics, PCM and Biology class and form a perception about how "chilly" the climate would be if they do choose these subjects post matriculation. If they see a higher share of their own gender among seniors who opted for the respective subjects, they may be more encouraged to opt for them. Out of Y seniors of a student who opted for a subject, if g are girls and b $(= Y - g)$ are boys, then

$$\begin{aligned} Chilly\ climate\ measure\ for\ girls &= \frac{g}{Y} \\ Chilly\ climate\ measure\ for\ boys &= \frac{b}{Y} \end{aligned} \tag{1}$$

The lower the value of this measure, the "chillier" the expected climate will be for a student.

Table 7 shows how the distribution of our measure of chilly climate varies across genders. We only look at co-educational schools for this part of the analysis where students have both male and female seniors. Column 2 reports the overall mean share of own gender students in senior Mathematics, PCM and Biology classes. On an average, a student sees that 55% of the seniors who opted for Mathematics or PCM are of her own gender, while 50% of Biology opting seniors are of her own gender. The next two columns report these values for girls and boys separately. Here, we see that while an average boy sees that almost 70% of the students in the senior Mathematics class are boys, an average girl sees that only 35% of those students are girls. The difference is even larger for PCM. The pattern expectedly flips for Biology with girls having 63% of own gender students in senior Biology class and boys having only 40%.

4.5 Socioeconomic characteristics

Finally, in this subsection, we describe the background socioeconomic characteristics of students. We already saw the statistics related to these variables in Table 1. Table 8 shows their distribution across girls and boys. More girls than boys belong to General caste category, while more boys belong to the OBC category. Fewer girls than boys are enrolled in Independent schools, more girls are enrolled in government schools. Finally, fewer girls than boys are single children in their families. The difference in the annual family income is not statistically significant due to the large standard deviations.

5 Decomposition Analysis

Our aim in this section is to formally decompose the gender gap in the rate of take-up of Mathematics, PCM and Biology in our dataset. For this, we use a regression framework and a decomposition framework. Below, we a give general description of both the methodologies.

5.1 Regression framework

We run the following linear probability model:

$$S_{isc} = \beta_0 + \beta_1.F_{isc} + \beta_2.EF_{isc} + C_c + \delta_d + \varepsilon_{isc}. \tag{2}$$

Here S_{isc} is the subject choice of student i in school s in cohort c. It is a binary variable which takes value 1 if the student chose subject S after class X, and 0 otherwise. F_{isc} is a dummy variable for a female student. EF_{isc} stands for Explanatory Factor measured at the level of a student i in school s in cohort c. It can be socioeconomic characteristics of students, a measure of their ability, variables related to cohort peers or those measuring the chilly climate aspect of immediate seniors. C_c are cohort fixed effects to control for unobservables within a cohort and δ_d are district fixed effects to control for unobservables at the district level.

In this framework, we will monitor the movement of the female dummy variable, F_{isc}. Without any controls, the coefficient of female dummy captures the raw difference between the take-up of a subject between boys and girls. We then add explanatory factors to see if it is sensitive to the addition of these controls.

5.2 Decomposition framework

Pioneered by Blinder (1973) and Oaxaca (1973) to study wage gaps, decomposition techniques "decompose" a gap in a distributional statistic between two groups into an "explained" and a residual, "unexplained" component (Fortin et al. 2011). The explained component is the one due to a difference in endowments between the two groups. The unexplained part is due to a difference in the returns to those endowments. In this paper, we use the Blinder-Oaxaca decomposition technique(Jann

2008).

It is commonplace in the decomposition literature to use the coefficients of the dominant group to calculate the explained part of the decomposition in the wage gap. The dominant group, in general, has higher returns to the given set of attributes like education and experience. This is not always the case in our context, as we will see in the next section. We instead use coefficients from a pooled model to determine the explained component as proposed by Neumark (1988) and Oaxaca and Ransom (1994). To avoid overestimation of the explained part due to inappropriate spillage of some of the unexplained component into the explained component, we also include a dummy for a female student as an additional covariate in the pooled model (Jann 2008). Let $E(X_B)$ be the mean level of an endowment for boys and $E(X_G)$ be that for girls. Then the explained component of the Blinder-Oaxaca decomposition is given by $\{[E(X_B) - E(X_G)]'\beta_P\}$, where β_P is the coefficient from the pooled regression and is assumed to be the same for girls and boys.

We now delve into each of the categories of explanatory factors one by one and report the results in the next section.

6 Results

6.1 Ability

We begin by reporting results from the linear probability models. Recall that we need to track the movement in the coefficient of the female dummy for our purposes. We add the socioeconomic status (SES) variables as the first set of explanatory variables. Thereafter we use them as controls with all other explanatory factors. Table 9 reports the first set of results. The first three columns have Mathematics as the subject choice, columns 4, 5 and 6 have PCM and the last three have Biology as the subject choice. The first columns for each subject only add the cohort and district fixed effects. The second columns add the SES variables, which include the (administrative) caste, annual family income and single child status of a student. The third columns for each subject add ability variables. These include the class X total score, class X Mathematics and Science scores and comparative advantage terms[24]. To recall, a student's comparative advantage in subject A vs. B is measured as the difference in her standardized scores in these two subjects.

For Mathematics, we saw from Table 2 that the raw gender gap is 19.12 percentage points. As explained in Section 5.1, this would be the value of the coefficient on the female dummy without any other variables on the right hand side. We see from column 1 of Table 9 that the addition of cohort and district fixed effects reduces the size of the coefficient on the female dummy by around 2 percentage points to 17.1 percentage points. The addition of SES variables in column 2 leaves the coefficient effectively unchanged. The addition various measures of absolute and comparative advantage in ability in column 3, on the other hand, actually increases the magnitude of the female dummy coefficient by 0.7 percentage points. Class X total score has a negative and statistically significant coefficient. While this may seem counterintuitive, we must note that the specification includes all the other ability terms and the *net* return to class X total score is negative. The size of the coefficient, however, is very small. The coefficients on both Mathematics score and Science score are positive but statistically significant only for Science score. Finally, a comparative ad-

[24]Since adding all three comparative advantage terms together leads to collinearity and one of the terms gets dropped, we include only two of them for each subject choice. A comparative advantage in Mathematics vs. English and in Science vs. English are added for Mathematics and PCM while a comparative advantage in Science vs. English and in Science vs. Mathematics are added for Biology.

vantage in Mathematics vs. English has a large, positive and statistically significant association with choosing Mathematics after class X, while a comparative advantage in Science vs. Mathematics has a negative and statistically significant correlation.

Columns 4 to 6 show very similar results for PCM. All terms, except the class X Mathematics score, have coefficients that are similar in magnitude, sign and statistical significance to those of Mathematics. The coefficient of the class X Mathematics score has a negative and statistically significant coefficient. Importantly, the addition of all ability terms together increases the magnitude of the female dummy coefficient for PCM too.

Columns 7 to 9 for Biology show that class X total score has a positive and statistically significant coefficient, as does the coefficient of class X Mathematics score. Class X Science score has a negative and statistically significant association with the take-up of Biology. Lastly, a comparative advantage in Science vs. Mathematics has a large and positive coefficient which is statistically significant. The coefficient of the Science vs. English comparative advantage term has a negative coefficient. Most importantly, while the fixed effects and SES variables increase the size of the female dummy coefficient above the raw gender gap of 11.6 percentage points observed in Table 2, the ability terms purge the size of the female dummy coefficient by 1 percentage point compared to the raw gap.

Table 10 reports results from the Blinder-Oaxaca decomposition. We measure the gap as take-up of boys minus that of girls. As explained in Section 5.2, the coefficients from Table 9 will be used to calculate the explained components. Columns 1 and 2 report results for Mathematics, followed by PCM in columns 3 and 4 and Biology in columns 5 and 6. The first row gives the total gap in the take-up in percentage points. The following rows give the contribution of SES and ability variables in percentage points. The numbers in the brackets below give the contribution as a percentage of the total gap ((percentage points explained/total gap)*100). The odd numbered columns only have the socioeconomic status variables. First consider Mathematics and PCM. The SES variables explain 0.79% of the total gap of 19.08 percentage points in Mathematics and a statistically insignificant portion of the 20.57 percentage points gap in PCM. The ability variables are added in the next columns. Class X total score explains less than 1% of the gap in Mathematics and in PCM. The contribution of the class X Mathematics score is statistically insignificant for both the subjects. The contribution of the class X Science score, on the other hand, is negative. To understand this mathematically, recall that the "explained" part of the gender gap is given by the term $\{[E(X_B) - E(X_G)]'\beta_P\}$, where β_P is the coefficient from the pooled regression of Table 9. Since girls have a higher mean Science score than boys, the expression inside the brackets is negative in sign, while β_P is positive, resulting in a negative product. Intuitively, this can be understood by using the counterfactual exercise. If girls had the mean Science scores of boys, which is lower than their actual scores, then the gender gap would, in fact, be higher than what is observed, since Science scores are positively correlated with the probability of choosing Mathematics.

The largest contribution to the explained part comes from the gender differences in comparative advantage in Mathematics vs English. It explains around 19% of the gap for both Mathematics and PCM. The Science vs. English comparative advantage term, however, has a negative contribution to the explained part. This is to be expected given the negative coefficient of this term in Table 9 and that girls have a higher mean value of this attribute. The last row shows that all the ability terms together explain 5.92% of the gender gap in Mathematics take-up and 4.42%

of that in PCM take-up.

For Biology, the total gap is -11.65 percentage points as the take-up rate of Biology is lower among boys compared to girls. Column 5 shows that when only SES variables are added, the explained percentage is statistically insignificant. Class X total score explains around 6% of the gap. Boys have lower class X total score on average, and if girls had scores like boys, they would be lesser likely to choose Biology and thus the gender gap will be smaller. The contribution of class X Mathematics score is statistically insignificant while that of the Science score is negative. This means the predicted gap is higher: if girls had Science scores like that of boys, they would have even higher probability of choosing Biology (The coefficient on Science score is negative in Table 9.). For Biology too, the largest contribution comes from the comparative advantage term. A comparative advantage in Science vs. Mathematics explains 9% of the gender gap in Biology. If girls were counterfactually given lower mean values of Science vs. Mathematics comparative advantage, they would be 9.01% less likely to choose Biology. All the ability factors together explain 10.64% of the 11.65 percentage points gender gap in Biology.

Summing up, comparative advantages in the relevant subjects turn out to be the largest explanators under the ability head. They explain around 19% of the gap in Mathematics and PCM and around 9% of the gap in Biology. Taken together, ability differences can explain less that 6% of the gap in Mathematics and PCM and a little over 10% of the gap in Biology.

Hereafter, we skip the results from the linear probability models for the remainder of the explanatory factors. The same are available from the authors upon request.

6.2 Cohort Peers

We now consider our next set of explanatory factors: cohort peers. As discussed in Section 4.3, we consider two aspects of the cohort peers of a student: the share of female students in her cohort, and the quality of these peers as measured by their average scores in class X board examinations.

Table 11 shows the decomposition results. Column 1 has Mathematics as the subject choice, column 2 has PCM and column 3 has Biology. The first row shows the total gender gap in this sample of students from co-educational schools. All specifications include the SES variables as controls.

For Mathematics, we observe that female share in cohort peers explains 3.65% of the gender gap in take-up. If girls had lower share of females in their cohorts as boys do, they would be 3.65% more likely to opt for Mathematics post matriculation. The highest contribution comes from average peer score. We saw in Table 6 that girls have poorer quality peers compared to boys. If they had higher performing peers on average, like boys do, they would be able to close the gender gap by around 8%. The contribution of average female peer score is statistically insignificant, while that of average male peer score is negative. Overall, the peer related attributes can explain 9.13% of the gender gap in Mathematics.

The results column 2 show that for PCM average cohort peer score is the only statistically significant contributor to the explained part. It can close the gender gap in PCM by around 5.5%. The total explained portion is also close to 5.5% since the contribution of the other peer related variables is statistically insignificant.

In case of Biology, column 3 shows that the individual and well as overall contributions of cohort peer related variables is either statistically insignificant or negative. Thus, if girls had the mean levels of female share in cohort and cohort peer scores like boys, they would be even more likely to opt for Biology, thus widening the gender gap in the take-up.

To sum up, among the variables related to cohort peers, average cohort peer score is the largest contributor to the explained part for Mathematics and PCM, while for Biology, the contribution of cohort peer variables is negative.

6.3 Immediate Seniors

We now move on to consider our last set of explanatory factors where we use a student's immediate seniors to elicit the chilly environment aspect of making a stream choice. As explained in Section 4.4, only the two younger cohorts can be used for this part of the analysis. The two older cohorts – class X batches of 2012 and 2013, are the immediate seniors of the two younger cohorts – class X batches of 2013 and 2014, respectively.

Table 12 shows decomposition results. Again, column 1 presents the results for Mathematics, column 2 for PCM and column 3 for Biology. We see that for Mathematics and PCM, the total explained gap is the largest so far: 13.90% and 15.82%, respectively. The individual contribution of our chilly climate measure is close to 13.5% for Mathematics and a little over 15% for PCM. Thus, if girls expected to face a "warm" climate upon choosing Mathematics and Biology like boys do, as is evident from Table 7, they would be 14% to 15% more likely to opt for these male dominated subjects.

Column 3 shows that of our chilly climate measure can explain around 6% of the gender gap in Biology. This contribution, while positive and statistically significant, is much smaller than that observed in columns 1 and 2. The total explained part is 5.35% for Biology.

Overall, our measure of chilly climate explains around 14% and 16% of the gender gap in Mathematics and PCM and around 5% of the gap in Biology. It is interesting to note that the expected chilliness in the climate of a course dominated by the opposite gender is the most important factor for girls for avoiding that course (compared to the other explanatory factors studied in this paper). It is also interesting that this is a much less important factor for boys. Thus, while a skewed sex ratio discourages girls to enter in Mathematics and PCM, this is not an important factor for boys when making a choice to take up female dominated courses like Biology.

6.4 All Factors

To directly compare their strength against each other, Table 13 presents decomposition results when all the factors considered so far are added together. Column 1 shows explained parts for Mathematics, column 2 for PCM and column 3 for Biology[25].

For Mathematics, we see that only our measure of chilly climate (share of own gender students in senior Mathematics class) has a positive and statistically significant contribution. It can explain around 13.5% of the gender gap in the take-up of the subject. The contribution of a comparative advantage in Mathematics vs. English, though large and positive, is statistically insignificant. Taken together, all the factors explain 9.09% of the gender gap in Mathematics.

For PCM too, the chilly climate measure is the largest contributor to the explained part (14.73%), closely followed by a comparative advantage in Mathematics vs. English (13.29%). Together, all the factors can explain around 8% of the gender gap in PCM.

On the other hand, for Biology, the largest contribution comes from class X total

[25]Due to the inclusion of variables related to peers and immediate seniors, the sample for this exercise is restricted to co-educational schools in the later two cohorts.

score which explains 12.08% of the gap. The chilly climate measure is the only other statistically significant contributor, explaining a little over 5% of the gender gap. All the variables together can explain over 16% of the gender gap in the take-up of Biology.

Thus, even when we put all the factors together, our observation from the previous subsection still holds. Our measure of chilly climate is the largest contributor to explaining the gender gap in the male dominated subjects of Mathematics and PCM, but not for a female dominated subject like Biology.

7 Discussion

While we are able to explore a wide range of factors possibly responsible for the stream choices of students, a caveat in our findings is that at least 84% of the gap remains unexplained even after accounting for all those factors[26]. In this section we discuss the possible sources for this residual gap and the scope of future research.

One of the major explanatory factors studied in the literature, which we were unable to examine given our dataset, is individual preference. A number of studies have shown that girls and boys have very different preferences regarding the non-pecuniary aspect of a career (Montmarquette et al. 2002; Baram-Tsabari and Yarden 2011; Kahn and Ginther 2017; Wang and Degol 2017; Patnaik et al. 2020). For example, girls place a higher value on enjoyability of the course material (Zafar 2013; Wiswall and Zafar 2015), people oriented careers (Diekman et al. 2010; Eccles and Wang 2016) and family-work life balance (Bronson 2014; Wasserman 2015) than boys. These differences could be a result of social conditioning where men and women are required to fill in preordained gender roles. Differential risk and confidence across

[26]The highest percentage of the gender gap is explained for Biology in Table 13, which is 16.30%.

genders could also shape their preferences. Finally, especially in the context of India, decision making agency is also crucial. Girls (and often boys) may not be choosing streams based on their preferences, but based on what their parents deem fit for them. The choice of parents could, in turn, be driven by marriage market demands, cost of education, willingness to spend on education as well as on availability of higher education institutes in their locality. For example, girls may not be allowed to take Mathematics or PCM because pursuing STEM programmes after high school either involves a higher cost of education, or a relevant college is not available in the vicinity and parents are reluctant to send their daughters to far away colleges. It could also be because a girl child is not expected to continue education after school. A boy, similarly, may be discouraged to opt for Biology because of social desirability bias, or even because a career in medicine has a long gestation period and boys are expected to start earning early on in their life.

The foremost implication of these is that the girl-boy difference in stream choice is not superficial. Rather, it is deep rooted in cultural upbringing and societal expectations (Giuliano 2020). The policies designed to bridge this gap have to incorporate these nuances and target deeper issues of gender stereotyping, marriage market functionalities, and post marriage balance of power and division of labour in the family. A most pressing future area of research would be to link career choices with marriage market variables.

8 Conclusion

In this paper we examine in detail the first stream choices of students in India after class X. Using three cohorts of results data from the Central Board of Secondary Education (CBSE), we first quantify the extent of the gender gap in subjects chosen in higher secondary school. We find that there is a 19.12 percentage points difference in the take-up Mathematics between boys and girls. The gap is 20.59 percentage points in PCM choice. Boys are also 5 percentage points more likely to opt for computer Science. Girls, on the other hand, are 11.60 percentage points more likely to choose Biology and 5 percentage points more likely to choose Arts and Economics.

Next we describe each category of our explanatory factors one by one and document how they are distributed differently across girls and boys. The largest differences between the genders are seen in the attributes related to their immediate seniors. Then we examine each explanatory factor under a decomposition framework using the Blinder-Oaxaca linear decomposition technique. We report three broad findings. One, the contribution of the chilly climate measure is asymmetric by whether a subject is male or female dominated. The chilly climate measure is the largest explanator of the gender gap in Mathematics and PCM, but not in Biology. If girls had the share of own gender students in the senior Mathematics and PCM classes like boys, the gender gap in these subjects would have closed by 14% and 16%, respectively. Two, for Mathematics and PCM, a comparative advantage in Mathematics vs. English is the second largest contributor to explaining the gap, while it is the class X total score for Biology. Three, when put together, peer composition and peer performance variables do not explain any statistically significant portion of the gender gap in any of the three subjects. This is possibly because the differences in peer attributes is small across genders. We also note that even after accounting for gender differences in a wide range of attributes, at least 84% of the gender gap in our data remains unexplained. The differences observed in the form of stream choices actually masks huge differences in upbringing, expectations and balance of power between men and women from a very young age.

References

Ajayi, K. F. and Buessing, M. (2015). Gender parity and schooling choices. *The Journal of Development Studies*, 51(5):503–522.

Anderson, J. and Lightfoot, A. (2019). The school education system in India: An Overview. *The British Council*.

Anelli, M. and Peri, G. (2019). The effects of high school peers' gender on college major, college performance and income. *The Economic Journal*, 129(618):553–602.

Ashraf, J. and Ashraf, B. (1993). Estimating the gender wage gap in Rawalpindi city. *The Journal of Development Studies*, 29(2):365–376.

Astorne-Figari, C. and Speer, J. D. (2019). Are changes of major major changes? The roles of grades, gender, and preferences in college major switching. *Economics of Education Review*, 70:75–93.

Baram-Tsabari, A. and Yarden, A. (2011). Quantifying the gender gap in science interests. *International Journal of Science and Mathematics Education*, 9(3):523–550.

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., Van Der Erve, L., Sibieta, L., Vignoles, A., Walker, I., and Zhu, Y. (2018). The relative labour market returns to different degrees: Research report. *Institute for Fiscal Studies*.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, Oct(1):436–455.

Boll, C. and Lagemann, A. (2018). Gender pay gap in EU countries based on SES (2014). *Luxembourg, Publication Office of the European Union. doi*, 10:978935.

Bostwick, V. K. and Weinberg, B. A. (2018). Nevertheless she persisted? Gender peer effects in doctoral STEM programs. *NBER Working Paper No. w25028*.

Bronson, M. A. (2014). Degrees are forever: Marriage, educational investment, and lifecycle labor decisions of men and women. *Unpublished manuscript*, 2.

Brown, C. J., Pagán, J. A., and Rodríguez-Oreggia, E. (1999). Occupational attainment and gender earnings differentials in Mexico. *ILR Review*, 53(1):123–135.

Buonanno, P. and Pozzoli, D. (2009). Early labour market returns to college subject. *Labour*, 23(4):559–588.

Chakrabarti, A. (2009). Determinants of participation in higher education and choice of disciplines: Evidence from urban and rural Indian youth. *South Asia Economic Journal*, 10(2):371–402.

Chanana, K. (2000). Treading the hallowed halls: Women in higher education in India. *Economic and Political Weekly*, pages 1012–1022.

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

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1):50–58.

Cheney, G. R., Ruzzi, B. B., and Muralidharan, K. (2005). A profile of the Indian education system. *Prepared for the New Commission on the Skills of the American Workforce*.

Chi, W. and Li, B. (2014). Trends in China's gender employment and pay gap: Estimating gender pay gaps with employment selection. *Journal of Comparative Economics*, 42(3):708–725.

Clark Blickenstaff, J. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and education*, 17(4):369–386.

Dahl, G. B., Rooth, D.-O., and Stenberg, A. (2020). Long-run returns to field of study in secondary school. *NBER Working Paper No. w27524*.

Das, P. (2012). Wage inequality in India: Decomposition by sector, gender and activity status. *Economic and Political Weekly*, pages 58–64.

Deshpande, A. (2011). *The grammar of caste: Economic discrimination in contemporary India*. Oxford University Press, New Delhi.

Deshpande, A., Goel, D., and Khanna, S. (2018). Bad karma or discrimination? Male–female wage gaps among salaried workers in India. *World Development*, 102:331–344.

Dickson, L. (2010). Race and gender differences in college major choice. *The Annals of the American Academy of Political and Social Science*, 627(1):108–124.

Diekman, A. B., Brown, E. R., Johnston, A. M., and Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological science*, 21(8):1051–1057.

Dolado, J. J., Felgueroso, F., and Jimeno, J. F. (2002). Recent trends in occupational segregation by gender: A look across the Atlantic.

Dolton, P. J. and Vignoles, A. (2002). The return on post–compulsory school mathematics study. *Economica*, 69(273):113–142.

Duraisamy, M. and Duraisamy, P. (2016). Gender wage gap across the wage distribution in different segments of the Indian labour market, 1983–2012: Exploring the glass ceiling or sticky floor phenomenon. *Applied Economics*, 48(43):4098–4111.

Eccles, J. S. and Wang, M.-T. (2016). What motivates females and males to pursue careers in mathematics and science? *International Journal of Behavioral Development*, 40(2):100–106.

England, P. and Li, S. (2006). Desegregation stalled: The changing gender composition of college majors, 1971-2002. *Gender & Society*, 20(5):657–677.

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.

Fletschner, D., Anderson, C. L., and Cullen, A. (2010). Are women as likely to take risks and compete? Behavioural findings from Central Vietnam. *The journal of development studies*, 46(8):1459–1479.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. In *Handbook of labor economics*, volume 4, pages 1–102. Elsevier.

Fouad, N. A., Singh, R., Fitzpatrick, M. E., and Liu, J. P. (2011). Stemming the tide: Why women leave engineering. *University of Wisconsin-Milwaukee, Final report from NSF Award*, 827553.

Friedman-Sokuler, N. and Justman, M. (2016). Gender streaming and prior achievement in high school science and mathematics. *Economics of Education Review*, 53:230–253.

Fryer Jr, R. G. and Levitt, S. D. (2009). An empirical analysis of the gender gap in Mathematics. *NBER Working Paper No. 15430*.

Gautam, M. (2015). Gender, subject choice and higher education in india: Exploring 'choices' and 'constraints' of women students. *Contemporary Education Dialogue*, 12(1):31–58.

Giuliano, P. (2020). Gender and culture. *NBER Working Paper No. w27725*.

Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5):1637–1664.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

Gneezy, U. and Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, 94(2):377–381.

Goulas, S., Griselda, S., and Megalokonomou, R. (2020). Comparative advantage and gender gap in STEM.

Grogger, J. and Eide, E. (1995). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources*, Apr(1):280–310.

Guimarães, C. R. F. F. and Silva, J. R. (2016). Pay gap by gender in the tourism industry of Brazil. *Tourism Management*, 52:440–450.

Hardies, K., Breesch, D., and Branson, J. (2013). Gender differences in overconfidence and risk taking: Do self-selection and socialization matter? *Economics Letters*, 118(3):442–444.

Hegewisch, A. and Hartmann, H. (2014). Occupational segregation and the gender wage gap: A job half done. *Cornell University, ILR School*.

Hegewisch, A., Liepmann, H., Hayes, J., and Hartmann, H. (2010). Separate and not equal? Gender segregation in the labor market and the gender wage gap. *IWPR Briefing Paper*, 377:1–16.

Hunt, J. (2016). Why do women leave science and engineering? *ILR Review*, 69(1):199–226.

Jakobsson, N. (2012). Gender and confidence: are women underconfident? *Applied Economics Letters*, 19(11):1057–1059.

James, E., Alsalam, N., Conaty, J. C., and To, D.-L. (1989). College quality and future earnings: Where should you send your child to college? *The American Economic Review*, 79(2):247–252.

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4):453–479.

Justman, M. and Méndez, S. J. (2018). Gendered choices of STEM subjects for matriculation are not driven by prior differences in mathematical achievement. *Economics of education review*, 64:282–297.

Kahn, S. and Ginther, D. (2017). Women and STEM. *NBER Working Paper No. w23525*.

Khanna, S. (2012). Gender wage discrimination in India: Glass ceiling or sticky floor? *Delhi School of Economics, Centre for Development Economics Working Paper*, (214).

Kugler, A. D., Tinsley, C. H., and Ukhaneva, O. (2017). Choice of majors: Are women really different from men? *NBER Working Paper No. w23735*.

Landaud, F., Ly, S. T., and Maurin, É. (2020). Competitive schools and the gender gap in the choice of field of study. *Journal of Human Resources*, 55(1):278–308.

Lordan, G. and Pischke, J.-S. (2016). Does Rosie like riveting? Male and female occupational choices. *NBER Working Paper No. w22495*.

Marsh, H. W., Martin, A. J., and Cheng, J. H. (2008). A multilevel perspective on gender in classroom motivation and climate: Potential benefits of male teachers for boys? *Journal of Educational Psychology*, 100(1):78.

Mattoo, M. I. (2013). Career choices of secondary students with special reference to gender, type of stream and parental education. *Research on Humanities and Social Sciences*, 3(20):55–61.

Maurer-Fazio, M., Rawski, T. G., and Zhang, W. (1999). Inequality in the rewards for holding up half the sky: Gender wage gaps in China's urban labour market, 1988-1994. *The China Journal*, (41):55–88.

McGuinness, S. (2003). University quality and labour market outcomes. *Applied Economics*, 35(18):1943–1955.

Menon, N. and Van der Meulen Rodgers, Y. (2009). International trade and the gender wage gap: New evidence from India's manufacturing sector. *World Development*, 37(5):965–981.

Montmarquette, C., Cannings, K., and Mahseredjian, S. (2002). How do young people choose college majors? *Economics of Education Review*, 21(6):543–556.

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, pages 279–295.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, pages 693–709.

Oaxaca, R. L. and Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1):5–21.

O'Neill, J. (2003). The gender gap in wages, circa 2000. *American Economic Review*, 93(2):309–314.

Park, G., Lubinski, D., and Benbow, C. P. (2007). Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science*, 18(11):948–952.

Patnaik, A., Wiswall, M. J., and Zafar, B. (2020). College majors. *NBER Working Paper No. w27645.*

Penner, A. M. and Paret, M. (2008). Gender differences in Mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, 37(1):239–253.

Pirinsky, C. (2013). Confidence and economic attitudes. *Journal of Economic Behavior & Organization*, 91:139–158.

Plantenga, J., Remery, C., et al. (2006). The gender pay gap. Origins and policy responses. A comparative review of thirty European countries. *European Commission, Directorate-General for Employment, Social Affairs and Equal Opportunities.*

Pope, D. and Sydnor, J. (2010). A new perspective on stereotypical gender differences in test scores. *Journal of Economic Perspectives*, 24(95):108.

Prakasam, G. R., Mukesh, and Gopinathan, R. (2019). Enrolment by academic discipline in higher education: Differential and determinants. *Journal of Asian Business and Economic Studies*, 26(2):265–285.

Rapoport, B. and Thibout, C. (2018). Why do boys and girls make different educational choices? The influence of expected earnings and test scores. *Economics of Education Review*, 62:205–229.

Ray, T., Chaudhuri, A. R., and Sahai, K. (2020). Whose education matters? An analysis of inter caste marriages in India. *Journal of Economic Behavior & Organization*, 176:619–633.

Reilly, B., Dutta, P. V., et al. (2005). The gender pay gap and trade liberalisation: Evidence for India. *Poverty Research Unit at Sussex Working Paper*, 32.

Rendall, M. (2013). Structural change in developing countries: Has it decreased gender inequality? *World Development*, 45:1–16.

Riegle-Crumb, C. and King, B. (2010). Questioning a white male advantage in STEM: Examining disparities in college major by gender and race/ethnicity. *Educational Researcher*, 39(9):656–664.

Riegle-Crumb, C., King, B., Grodsky, E., and Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal*, 49(6):1048–1073.

Rubery, J. (1992). Pay, gender and the social dimension to Europe. *British Journal of Industrial Relations*, 30(4):605–621.

Sadker, M. and Sadker, D. (1986). Sexism in the classroom: From grade school to graduate school. *The Phi Delta Kappan*, 67(7):512–515.

Sahoo, S. and Klasen, S. (2020). Gender segregation in education and its implications for labour market outcomes: Evidence from India. *Demography (forthcoming).*

Sarsons, H. and Xu, G. (2015). Confidence men? Gender and confidence: Evidence among top economists. *Harvard University, Department of Economics, Littauer Center*, pages 1–26.

Tellhed, U., Bäckström, M., and Björklund, F. (2017). Will I fit in and do well? The importance of social belongingness and self-efficacy for explaining gender differences in interest in STEM and HEED majors. *Sex Roles*, 77(1-2):86–96.

Turner, S. E. and Bowen, W. G. (1999). Choice of major: The changing (unchanging) gender gap. *ILR Review*, 52(2):289–313.

Valla, J. M. and Ceci, S. J. (2014). Breadth-based models of women's underrepresentation in stem fields: An integrative commentary on Schmidt (2011) and Nye et al.(2012). *Perspectives on Psychological Science*, 9(2):219–224.

Wai, J., Cacchio, M., Putallaz, M., and Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence*, 38(4):412–423.

Wang, M.-T. and Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review*, 29(1):119–140.

Wasserman, M. (2015). Hours constraints, occupational choice and fertility: Evidence from medical residents. *MIT Working Paper.*

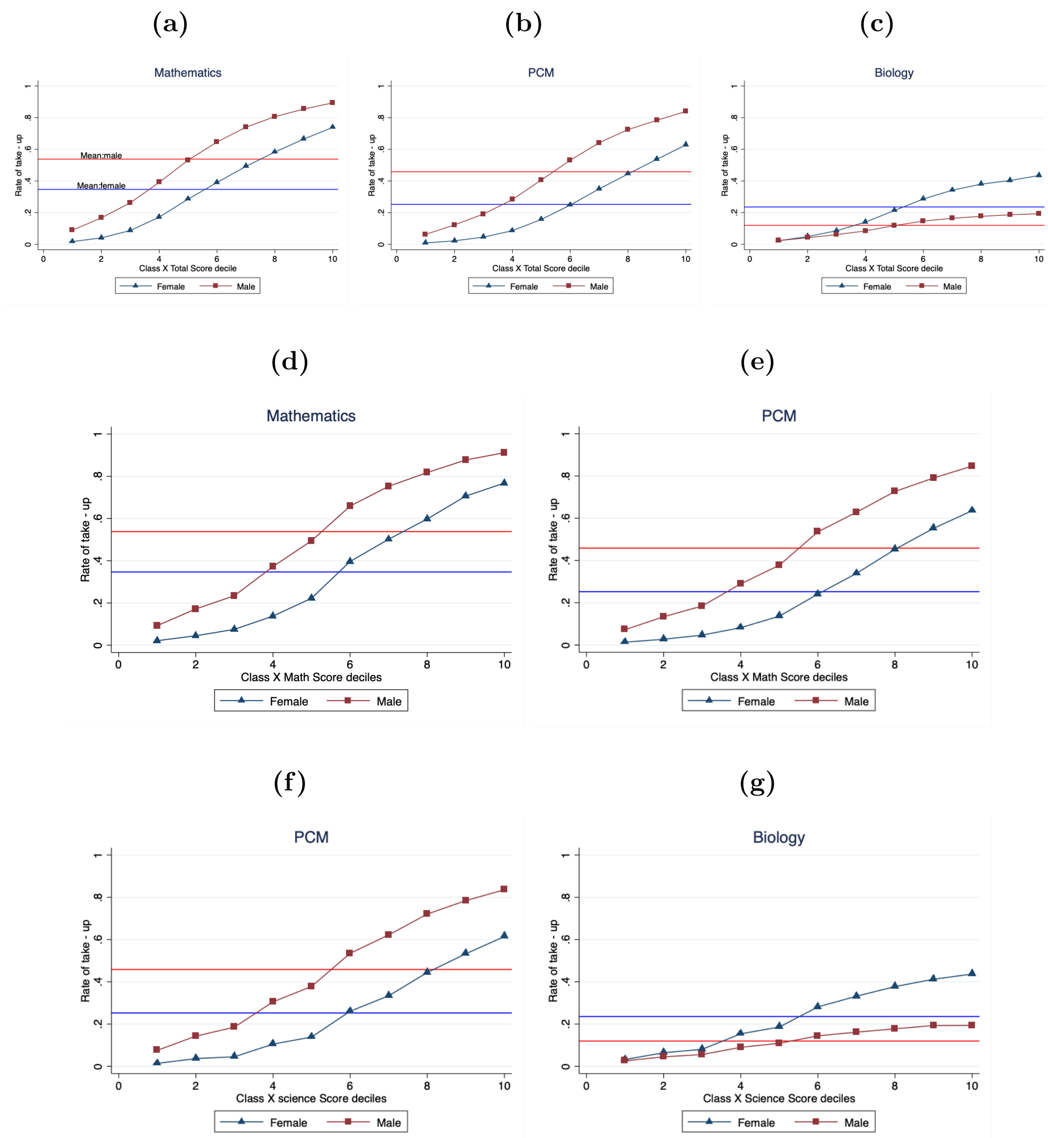
Webber, D. A. (2016). Are college costs worth it? How ability, major, and debt affect the returns to schooling. *Economics of Education Review*, 53:296–310.

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

Wu, A. H. (2017). Gender stereotyping in academia: Evidence from economics job market rumors forum. *Princeton University, Woodrow Wilson School of Public and International Affairs, Center for Health and Wellbeing, Working Paper 2017-09.*

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595.

Figure 1: Stream choice by class X scores



Note: Each graph plots the gender wise proportion of students who opt for a given subject after matriculation over class X score deciles. The first row plots the take-up of Mathematics, PCM and Biology over class X total score deciles. The second row plots the take-up of Mathematics and PCM over class X Mathematics score deciles. The third row plots the take-up of PCM and Biology over class X Science score deciles. Data source: Central Board of Secondary Education.

Table 1: Summary Statistics

Total students	2,405,349
Girls	43.99%
Mean age in class X	16.68 years
Caste	
General	73.06%
Scheduled Castes	7.31%
Scheduled Tribes	3.32%
Other Backward Castes	16.31%
Type of school administration (Class XII)	
Private Aided	1.68%
Government	19.55%
Independent	68.04%
Jawahar Navodaya Vidyalaya	3.04%
Kendriya Vidyalaya	7.40%
Mean score (out of 500) (std. dev)	
Class X	348.91
	(73.96)
Class XII	329.93
	(83.67)
Other attributes	
Mean Annual family income (INR)	292,655.8
(std. dev)	(1,757,698)
Single child	5.82%

Note: Castes are the administrative caste categories in India. Data source: Central Board of Secondary Education.

Table 2: Subject choice

Subject	Share of students (percent)	Of which girls (percent)	Share of girls (percent)	Share of boys (percent)	Difference (Boy – girl) (in percentage points)
Mathematics	45.40	33.62	34.70	53.81	19.12***
Physics	48.68	37.58	41.60	54.25	12.65***
Chemistry	48.89	37.61	41.80	54.46	12.66***
Biology	17.06	60.75	23.56	11.95	-11.60***
Computer Science	9.65	30.16	6.62	12.03	5.41***
History	15.58	54.87	19.44	12.56	-6.88***
Political Science	17.14	55.51	21.63	13.61	-8.02***
Geography	8.26	42.72	8.02	8.45	0.43***
Economics	35.17	47.56	38.03	32.93	-5.11***
Hindi	16.36	56.02	20.84	12.85	-7.99***
English	96.49	43.19	94.74	97.87	3.13***
Business Studies	29.57	46.01	30.93	28.50	-2.43***
Accounts	29.60	45.89	30.88	28.60	-2.28***
Subject combinations					
PCM					
(Phy+Chem+Math)	36.79	30.20	25.25	45.84	20.59***
Arts					
(History+Pol. Sc.)	13.01	54.07	15.99	10.67	-5.32***
Commerce					
(Bus. Studies+Acc)	29.17	46.0	30.52	28.11	-2.41***

Note: The table reports the pattern of stream choice of the most frequently offered subjects by CBSE schools post matriculation. The first column shows the subject/subject combination. The second column shows the percentage of all students who opt for that subject. The third column shows what percentage of students who opt for the subject are girls. The next two columns show the percentage of girls and boys, respectively, who opt for that subject. The last column reports differences between columns 5 and 4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data source: Central Board of Secondary Education.

Table 3: Gender wise class X scores

Subject	Overall mean (std. dev)	Girls' mean (std. dev)	Boys' mean (std. dev)	Difference (Boy – girl) (in std. dev terms)
English	70.51	71.40	69.81	-0.11***
	(14.83)	(15.08)	(14.59)	
Hindi	73.33	74.63	72.31	-0.17***
	(13.80)	(13.80)	(13.72)	
Mathematics	66.41	66.02	66.73	0.04***
	(17.45)	(17.35)	(17.52)	
Science	67.43	67.70	67.21	-0.03***
	(16.39)	(16.36)	(16.42)	
Social science	69.55	70.16	69.06	-0.07***
	(16.12)	(16.20)	(16.04)	
Total	348.91	351.66	346.75	-0.07***
	(73.96)	(74.24)	(73.67)	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Significance levels in the last column indicate the statistical significance of the differences between columns 4 and 3. Standard deviations in parenthesis. Data source: Central Board of Secondary Education.

Table 4: Gender wise comparative advantage

	Girls	Boys	Difference (Boy – girl)
Mean standardized scores			
Mathematics	-0.023	0.018	0.04***
English	0.060	-0.047	-0.11***
Science	0.017	-0.013	-0.03***
Mean gap in scores			
Mathematics-English	-0.081	0.066	0.15***
Science-English	-0.041	0.035	0.08***
Mathematics-Science	-0.040	0.031	0.07***
Percent who score better in (%)			
Mathematics vs English	46.67	56.43	
Science vs English	49.42	55.19	
Mathematics vs Science	49.71	55.82	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Significance levels in the last column indicate the statistical significance of the differences between columns 3 and 2. Scores are standardized by subtracting the cohort average from a student's own score and dividing the difference by the standard deviation of scores in the cohort. Data source: Central Board of Secondary Education.

Table 5: Gender composition and school type

Panel A			
Type of school	**Number of schools**	**Number of students**	**Average female share in cohort (%) (std.dev)**
Total	12,691	24,05,349	43.99
			(23.75)
All girls	1046	2,18,048	1
			(0)
All boys	1347	1,64,234	0
			(0)
Co-education	11,253	20,23,067	41.52
			(13.01)

Panel B	
Group	**Average female share in cohort (%) (std.dev)**
Girls	56.81
	(25.61)
Boys	33.92
	(16.16)
Girls in co-ed schools	45.60
	(14.71)
Boys in co-ed schools	38.63
	(10.75)

Note: Panel A gives summary statistics of different types of schools based on gender composition. Panel B reports gender wise peer composition of the cohort of an average student. Data source: Central Board of Secondary Education.

Table 6: Gender wise peer performance

Mean peer performance (out of 500)	**Overall (std. dev)**	**For girls (std. dev)**	**For boys (std. dev)**	**Difference (Boy – girl)**
Panel A: All schools				
All peers	348.90	345.15	351.84	6.69***
	(48.84)	(49.97)	(47.72)	
Female peers	359.98	351.61	367.45	15.84***
	(40.03)	(52.29)	(44.60)	
Male peers	347.21	347.99	346.73	-1.26***
	(48.40)	(48.76)	(48.16)	
Panel B: Co-educational schools				
All peers	356.87	354.35	358.66	4.31***
	(45.59)	(47.61)	(44.02)	
Female peers	365.40	362.50	367.45	4.95***
	(46.54)	(49.00)	(44.60)	
Male peers	350.83	347.99	352.85	4.86***
	(46.89)	(48.74)	(45.40)	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Significance levels in the last column indicate the statistical significance of the differences between columns 4 and 3. Mean peer performance is calculated by subtracting a student's own score from the sum total of scores of the students in her cohort in her school, and dividing the difference by the number of students in her school cohort minus one. Data source: Central Board of Secondary Education.

Table 7: Gender composition of senior class

Chilly climate measure	Overall mean (%) (std. dev)	Mean for girls (%) (std. dev)	Mean for boys (%) (std. dev)	Difference (Boy – girl)
Own gender in senior Math class	55.05	34.45	69.45	35.00***
	(23.20)	(16.81)	(14.58)	
Own gender in senior PCM class	55.43	31.59	72.06	40.47***
	(25.51)	(17.41)	(14.85)	
Own gender in senior Bio class	49.41	63.30	39.74	-23.56***
	(23.84)	(20.66)	(20.94)	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Significance levels in the last column indicate the statistical significance of the differences between columns 4 and 3. The table shows the gender composition of students' seniors' classes. Data source: Central Board of Secondary Education.

Table 8: Gender wise socioeconomic characteristics

Attribute	Mean for girls (%)	Mean for boys (%)	Difference (Boy – girl)
Caste			
General	74.45	71.96	−2.49***
Scheduled Castes	7.70	7.00	−0.70***
Scheduled Tribes	3.60	3.10	−0.50***
Other Backward Classes	14.25	17.93	3.68***
Type of school administration			
Private aided	1.85	1.54	−0.31***
Government	24.83	15.41	−9.42***
Independent	62.37	72.49	10.12***
Jawahar Navodaya Vidyalaya	2.69	3.32	0.63***
Kendriya Vidyalaya	8.04	6.90	−1.14***
Type of board examination in class X			
External board exam	35.88	38.82	2.94***
School board exam	64.12	61.18	−2.94***
Other attributes			
Mean Annual family income (INR)	293,174	292,248.9	-925
(std. dev)	(6,96,698.5)	(22,48,034)	
Single child	5.26	6.25	0.99***

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Significance levels in the last column indicate the statistical significance of the differences between columns 3 and 2. Castes are the administrative caste categories of India. Data source: Central Board of Secondary Education.

Table 9: Stream choice and ability

	(1) math	(2) math	(3) math	(4) pcm	(5) pcm	(6) pcm	(7) bio	(8) bio	(9) bio
F	-0.171***	-0.170***	-0.177***	-0.179***	-0.178***	-0.185***	0.125***	0.126***	0.106***
	(0.00247)	(0.00242)	(0.00182)	(0.00196)	(0.00193)	(0.00167)	(0.00158)	(0.00157)	(0.00141)
Class X total score			-0.000181***			-0.000350***			0.00141***
			(0.0000615)			(0.0000523)			(0.0000386)
Class X Math score			0.000463			-0.00497**			0.00296*
			(0.00264)			(0.00227)			(0.00177)
Class X Science score			0.0169***			0.0207***			-0.00477**
			(0.00279)			(0.00240)			(0.00188)
CA in Math vs Eng			0.248***			0.267***			
			(0.0456)			(0.0393)			
CA in Science vs Eng			-0.243***			-0.253***			-0.00220
			(0.0455)			(0.0392)			(0.00136)
CA in Science vs Math									0.149***
									(0.0308)
SES		√	√		√	√		√	√
Cohort FE	√	√	√	√	√	√	√	√	√
District FE	√	√	√	√	√	√	√	√	√
N	2384412	2384412	2384371	2384371	2384371	2384371	2384371	2384371	2384371
R^2	0.124	0.128	0.369	0.169	0.172	0.350	0.120	0.121	0.165

Note: Linear probability results are reported. Outcome is a dummy variable which takes value 1 if subject is chosen. Robust standard errors clustered at the school level are in parenthesis. CA stands of Comparative advantage, calculated by subtracting the standardized scores of one subject from another. SES stands for socioeconomic status variables which include caste dummies, annual family income and single child status of the student. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data source: Central Board of Secondary Education.

Table 10: Gender gap decomposition: Ability

Blinder-Oaxaca decomposition						
Explanatory factor added	**Explained (percentage points)** **Percent explained (%)**					
	(1) Mathematics	(2) Mathematics	(3) PCM	(4) PCM	(5) Biology	(6) Biology
Total gap (percentage points)	19.08	19.08	20.57	20.57	-11.65	-11.65
SES	0.15***	0.08***	0.16	0.14***	0.09	0.09***
	(0.79%)	(0.42%)	(0.78%)	(0.68%)	(-0.77%)	(-0.77%)
Class X total score		0.09**		0.18***		-0.71***
		(0.47%)		(0.88%)		(6.09%)
Class X Math score		0.03		-0.34		0.20
		(0.16%)		(-1.65%)		(-1.72%)
Class X Science score		-0.86**		-1.06**		0.24*
		(-4.51%)		(-5.15%)		(-2.06%)
CA in Math vs Eng		3.64***		3.91***		
		(19.08%)		(19.01%)		
CA in Science vs Eng		-1.84***		-1.91***		-0.02
		(-9.64%)		(-9.28%)		(0.17%)
CA in Science vs Math						-1.05***
						(9.01%)
Total	0.14***	1.13***	0.15	0.91***	0.09	-1.24***
	(0.73%)	(5.92%)	(0.73%)	(4.42%)	(-0.77%)	(10.64%)

Note: Decomposition results are reported. The first row shows the gender gap in take-up in percentage points for the column. The following rows report the detailed decomposition contribution of ability related variables in percentage points. The terms in the brackets report the contribution as a percent of the total gap shown in the top row. Robust standard errors clustered at the school level are in parenthesis. CA stands for Comparative Advantage, measured as the difference in class X standardized scores of the two subjects. SES stands for socioeconomic status variables which include caste dummies, annual family income and single child status of the student. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data source: Central Board of Secondary Education.

Table 11: Gender gap decomposition: Cohort peers

Blinder-Oaxaca decomposition			
Explanatory factor added	**Explained (percentage points) Percent explained (%)**		
	(1) Mathematics	(2) PCM	(3) Biology
Total gap (percentage points)	19.16	20.64	-14.90
Female share in cohort	0.70***	0.23	1.05***
	(3.65%)	(1.11%)	(-7.05%)
Average peer score	1.54***	1.13***	0.58**
	(8.04%)	(5.48%)	(-3.89%)
Average female peer score	-0.09	-0.15	-0.05
	(-0.47%)	(-0.73%)	(0.34%)
Average male peer score	-0.46***	-0.21	-0.21*
	(-2.40%)	(-1.02%)	(1.41%)
SES	0.10***	0.17***	0.07***
	(0.52%)	(0.82%)	(-0.47%)
Total	1.75***	1.14***	1.48***
	(9.13%)	(5.52%)	(-9.93%)

Note: Decomposition results are reported. The first row shows the gender gap in take-up in percentage points for the column. The following rows report the detailed decomposition contribution of cohort-peer related variables in percentage points. The terms in the brackets report the contribution as a percent of the total gap shown in the top row. Robust standard errors clustered at the school level are in parenthesis. Sample is restricted to co-educational schools. SES stands for socioeconomic status variables which include caste dummies, annual family income and single child status of the student. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data source: Central Board of Secondary Education.

Table 12: Gender gap decomposition: Immediate seniors

Blinder-Oaxaca decomposition			
Explanatory factor added	**Explained (percentage points) Percent explained (%)**		
	(1) Mathematics	(2) PCM	(3) Biology
Total gap (percentage points)	18.70	20.16	-15.89
Percent of own gender in senior subject class	2.52***	3.04***	-0.93***
	(13.48%)	(15.08%)	(5.85%)
SES	0.08**	0.15***	0.08***
	(0.43%)	(0.74%)	(-0.50%)
Total	2.60***	3.19***	-0.85***
	(13.90%)	(15.82%)	(5.35%)

Note: Decomposition results are reported. The first row shows the gender gap in take-up in percentage points for the column. The following rows report the detailed decomposition contribution of variables related to immediate seniors in percentage points. The terms in the brackets report the contribution as a percent of the total gap shown in the top row. Robust standard errors clustered at the school level are in parenthesis. Sample is restricted to co-educational schools in the later two cohorts of 2015 and 2016. SES stands for socioeconomic status variables which include caste dummies, annual family income and single child status of the student. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data source: Central Board of Secondary Education.

Table 13: Gender gap decomposition: All factors

Explanatory factor added	**Blinder-Oaxaca decomposition Explained (percentage points) Percent explained (%)**		
	(1) **Mathematics**	(2) **PCM**	(3) **Biology**
Total gap (percentage points)	18.70	20.16	-15.89
SES	0.05***	0.12***	0.07***
	(0.27%)	(0.60%)	(-0.44%)
Class X total score	-0.18*	-0.22*	-1.92***
	(0.96%)	(-1.09%)	(12.08%)
Class X Math score	-0.30*	-0.10	0.07
	(-1.60%)	(-0.50%)	(-0.44%)
Class X Science score	-1.21*	-1.95**	0.12
	(-6.47%)	(-9.67%)	(-0.76%)
CA in Math vs Eng	2.17	2.68**	
	(11.60%)	(13.29%)	
CA in Science vs Eng	-1.14	-1.46*	0.09**
	(-6.10%)	(-7.24%)	(-0.57%)
CA in Science vs Math			-0.64
			(4.03%)
Female share in cohort	0.15	-0.08	0.56***
	(0.80%)	(-0.40%)	(-3.52%)
Average peer score	0.19	-0.00	0.03
	(1.02%)	(-0.00%)	(-0.19%)
Average female peer score	-0.14*	-0.16**	-0.03
	(-0.75%)	(-0.79%)	(0.19%)
Average male peer score	-0.40***	-0.20*	-0.11
	(-2.14%)	(-0.99%)	(0.69%)
Percent of own gender in senior subject class	2.52***	2.97***	-0.83***
	(13.48%)	(14.73%)	(5.22%)
Total	1.70***	1.61***	-2.59***
	(9.09%)	(7.99%)	(16.30%)

Note: Decomposition results are reported. The first row shows the gender gap in take-up in percentage points for the column. The following rows report the detailed decomposition contribution of variables related to ability, classroom peers and immediate seniors in percentage points. The terms in the brackets report the contribution as a percent of the total gap shown in the top row. Robust standard errors clustered at the school level are in parenthesis. Sample is restricted to co-educational schools in the later two cohorts of 2015 and 2016. SES stands for socioeconomic status variables which include caste dummies, annual family income and single child status of the student. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data source: Central Board of Secondary Education.