

Kumar, Anand; Sahoo, Soham

**Working Paper**

## Social Identity and STEM Choice: Evidence from Higher Secondary Schooling in India

GLO Discussion Paper, No. 900

**Provided in Cooperation with:**

Global Labor Organization (GLO)

*Suggested Citation:* Kumar, Anand; Sahoo, Soham (2021) : Social Identity and STEM Choice: Evidence from Higher Secondary Schooling in India, GLO Discussion Paper, No. 900, Global Labor Organization (GLO), Essen

This Version is available at:

<https://hdl.handle.net/10419/235971>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# **Social Identity and STEM Choice: Evidence from Higher Secondary Schooling in India**

Anand Kumar<sup>1</sup> and Soham Sahoo<sup>2</sup>

July 2021

## **Abstract**

This paper investigates the role of social identity, namely gender and caste of individuals, in stream choice at the higher secondary level of schooling in India. We analyze the choice of science stream that is a crucial determinant of subsequent science, technology, engineering, and mathematics (STEM) related education and labour market outcomes. Using three rounds of nationally representative surveys between 2007 and 2018, we estimate regression models including household characteristics and fixed effects to control for region-specific unobserved heterogeneity. We find that females and individuals belonging to historically disadvantaged social groups (castes) are significantly less likely to study science. On average, female students enrolled in higher secondary education have 9 percentage points lower probability than male students to study science during the period considered for the analysis. For students belonging to the socially disadvantaged groups – Scheduled Tribes (ST), Scheduled Caste (SC), and Other Backward Classes (OBC), the likelihood of studying science is significantly lower by 6.9, 4.4, and 2.7 percentage points, respectively, than students from the less-disadvantaged groups. Using decomposition analysis, we show that the differences in the observable characteristics such as economic affluence and household head's education significantly explain some portion of the caste-based gaps, but they do not explain the gender gap that persists even in an intra-household comparison. We also test for the intersectionality of the overlapping identities of gender and caste in this context. We find evidence of cumulative disadvantages for females of SC and OBC groups, but lower gender disparity among ST students indicating the presence of intersectionality for this group. Thus, our paper highlights social identity as an important determinant of STEM-related education that may have long-term implications for the economic welfare of disadvantaged groups.

**Keywords:** Post-secondary education; STEM; Gender; Caste; Intersectionality; India

**JEL Codes:** I24; J15; J16

---

<sup>1</sup> Indian Institute of Management Bangalore. Email: [anand.kumar18@iimb.ac.in](mailto:anand.kumar18@iimb.ac.in)

<sup>2</sup> Indian Institute of Management Bangalore. Email: [soham.sahoo@iimb.ac.in](mailto:soham.sahoo@iimb.ac.in)

## 1. Introduction

Science, technology, engineering, and mathematics (STEM) education, in its role of developing the necessary knowledge and skills for innovative technological solutions, holds the key to Sustainable Development Goals (SDG) of building inclusive and sustainable societies (UNCTAD, 2018). Globally, women are less likely to pursue STEM majors and careers, with gender differences in stream choices emerging by secondary education when specialization begins (UNESCO, 2017). Similarly, the disparities in STEM education by family socio-economic status and race have been persistent (Xie et al., 2015). The under-representation of women and disadvantaged socio-economic groups in STEM fields is suboptimal, as the lack of diversity may attenuate collective understanding and problem-solving capabilities (Hong & Page, 2004). It also indicates that currently, STEM fields are unable to draw talents from the widest pool. The role of social identities (such as gender, race, ethnicity, or socio-economic group of the family), over which individuals do not have control, in deciding the academic and career choices in STEM fields, is testimony to the persisting inequalities of opportunities in education (Roemer & Trannoy, 2015).

The importance of technological progress in economic growth and development raises concerns regarding the missed opportunities of the under-represented groups. Also, at an individual level, gender and socio-economic segregations have implications for socio-economic mobility and welfare. While the STEM gaps for females and the marginalized socio-economic groups are widely documented at the university level (Gemici & Wiswall, 2014; Wiswall & Zafar, 2018; Zafar, 2013), the academic track trajectories are set at a much earlier stage in the schooling years. There has been a global movement in the policy space towards addressing the STEM gaps in the developed nations (Kearney, 2011; White House Report, 2013). However, in developing nations, the policy attention in addressing the gaps or documenting the gaps has been wanting.

In this paper, we study the effects of social identity on stream (specifically science stream) choice among students in the higher secondary level of education in India. These stream choices are important career decisions as they have been found to significantly affect the labour market outcomes later such as labour force participation and earnings (Jain et al., 2018; Roychowdhury, 2021; Sahoo & Klasen, 2018). In particular, we consider the social identities in terms of gender and caste to explore their relationship with choosing science at the higher

secondary level.<sup>3</sup> Firstly, we investigate the differences in the science take-up rates between the groups marked by the social identities of gender and caste. In the second stage of analysis, we decompose these gaps in science choice into parts that can be explained by the differences in the observable characteristics between the groups and parts that cannot be explained. With this approach, we seek to distinguish the nature of gaps in science choice between groups defined by gender and caste identities. Finally, we delve into the interaction of the two dimensions of social identity – gender and caste – using intersectionality analysis (Greenman & Xie, 2008). We investigate whether the gap in science choice among groups with overlapping identities is a summation of gaps of the constituting identities. Specifically, we test if the disadvantages of gender and caste add up for the female students of disadvantaged caste groups. This additive disadvantage of groups with multiple subordinate identities is called “double jeopardy” in the literature on race and gender (Crenshaw, 1990; King, 1988). This additive disadvantage notion is based on a strong assumption that the gender gaps across the caste groups are the same; and similarly, the caste gaps are the same across both genders. Conversely, there could be a positive (or a negative) interaction so that the effect of gender and caste is less (more) than their sum. In this case, the gender gap may vary significantly across different caste groups and vice versa; this phenomenon would indicate the presence of intersectionality in our context. We test these competing hypotheses for the interaction of gender and caste gaps in the context of science choice.

We use three rounds of nationally representative household surveys on education conducted in 2007-08, 2014, and 2017-18. Analyzing the repeated cross-sectional data spanning a decade, we find that females and disadvantaged castes have a significantly lower probability of attending the science stream at the higher secondary level of schooling. The estimates reveal a pro-male gender gap of about 9 percentage points among students enrolled at this level. For the socially disadvantaged groups – Scheduled Tribes (ST), Scheduled Castes (SC), and Other Backward Classes (OBC) – we find that the science take-up rates are significantly lower by 6.9, 4.4, and 2.7 percentage points, respectively, relative to the students from non-disadvantaged other castes (OC). Our regression model controls for various household characteristics such as household head’s education level and monthly per capita consumption expenditure as a proxy for the economic affluence of the household; thus, the estimated effect of caste is likely to reflect the influence of social identity rather than differences due to

---

<sup>3</sup> Castes, in India, are a system of rigid hierarchical social groups, which affect the social and economic lives of its members (Deshpande, 2011; Mosse, 2018; Munshi, 2019).

economic class or wealth. Further, we show that the estimated gender gap remains unchanged even when we include household fixed effects in the regression to compare boys and girls within the same household. We also find that these gender-based and caste-based gaps in science take-up have persisted over the decade with no significant change over the years. These results remain robust across various regression specifications including fixed effects to control for unobserved regional heterogeneity and addressing the potential sample selection due to endogenous enrollment in higher secondary education.

We further conduct a decomposition analysis to understand the drivers of the gaps. The decomposition exercise reveals an important distinction between the gender-based and caste-based gaps in science choice. A substantial part of the gap in science take-up rates between the caste groups is significantly explained by the differences in individual and household level characteristics. On the contrary, the explained part of the gender gap is insignificant. This finding implies that the disparity in science choice observed between caste groups is partially attributable to the socio-economic differences between the caste groups, but the same explanation is not applicable for the gap between male and female students. This result indicates that the gender gap in science choice is likely to be affected by unobserved factors such as gender norms, preferences, and decisions about gender-specific educational investment within the household.<sup>4</sup>

After analyzing the caste and gender gaps separately, we test the intersectionality versus double jeopardy hypothesis considering the interaction between these two dimensions of social identity. We find that the double jeopardy hypothesis holds for the female students in the OBC and SC groups, for whom the gender and caste disadvantages add up and yield a cumulative disadvantage. However, the gender gap is significantly lower among ST students than students from the non-disadvantaged caste, supporting the existence of intersectionality for the female students of the ST group.

Our study contributes to the literature in multiple ways. Firstly, we document the gender- and caste-based gaps in science choice using large-scale nationally representative data over a decade in the context of India. To the best of our knowledge, this is the first study documenting the caste-based gaps in stream choice at the higher secondary level of schooling in India.<sup>5</sup>

---

<sup>4</sup> Our finding is consistent with the literature showing the persistence of intra-household gender inequality in various educational outcomes (Kingdon, 2002; Sahoo, 2017; Sahoo & Klasen, 2018).

<sup>5</sup> On the gender gap in science choice, the study closest to our paper is Sahoo & Klasen (2021) which documents the gender gap in science and commerce choice using the India Human Development Survey

Secondly, we contribute to the understanding of how identities in separate dimensions, namely gender, and caste, influence education decisions differentially. Thirdly, we contribute to the literature on double jeopardy and intersectionality analysis in the context of India. The literature on caste and gender intersectionality has focused on the inequalities in educational attainment, nutrition, health, and household ownership (Deshpande, 2007); participation in market work (Eswaran et al., 2013); and wages (Agrawal, 2014). We contribute to this strand of literature in the context of STEM education through science stream choice at the higher secondary level.

The remainder of the paper is organized as follows: Section 2 discusses the background and related literature. Section 3 describes the data and summary statistics, Section 4 describes the empirical methodology, Section 5 presents the results and Section 6 concludes.

## **2. Background and Literature**

### **2.1 Stream Choice in the Indian Context**

In the Indian education system, a student undergoes 10 years of schooling following a non-selective curriculum. This includes the primary, middle and secondary levels of schooling with the structure being standard across the country. After completion of secondary schooling, the students enroll into two years of the higher secondary level, where they choose to specialize in mainly one of the following streams: science, commerce, or humanities, with a small proportion of students enrolling in other vocational tracks. After higher secondary schooling, they go onto tertiary education for Bachelor's and Master's programs in a chosen stream. Thus, in the Indian context, the first stage of stream selection happens at the higher secondary level. This choice largely determines the stream choices later at the tertiary level of education. Specifically, only the students who have specialized in the science stream at the higher secondary level are deemed eligible to pursue STEM-related courses at the tertiary level. This excludes the students who have specialized in commerce or humanities from taking up STEM courses. Thus, stream choice at the higher secondary level becomes a key decision that affects students' stream choices at subsequent levels of education which consequently affects the labour market outcomes later in their adult life.

---

of 2011-12. In this paper, we use larger and more updated surveys spanning the recent decade to analyze the trends in both gender and caste gaps.

Much of the literature on post-secondary stream choice comes from studies based in developed countries. Some of these studies have shown that interest in STEM subjects through the schooling years can predict students opting for STEM fields in later stages (Maltese et al., 2014; Sadler et al., 2012). Also, the labour market outcomes in terms of the expected lifetime earning potential from different streams could influence the stream choice (Altonji et al., 2016; Beffy et al., 2012; Boudarbat & Montmarquette, 2009). In the Indian setup, urban males who studied science at the higher secondary level have higher earnings relative to those who studied commerce or humanities (Jain et al., 2018; Roychowdhury, 2021). Sahoo & Klasen (2018) find that women who specialize in science stream at the higher secondary level have a higher probability of participating in the labour force, better occupations, and higher earnings.

In India, the Information Technology and Business Process Management (IT-BPM) industry has seen consistent growth and has been the flag bearer of India's exports in the last 20 years (Economic Survey, 2021). The growth of the IT-BPM industry has been shown to affect education participation and human capital investments (Jensen, 2012; Oster & Steinberg, 2013; Shastry, 2012). Despite India contributing nearly one-third (29.2 percent) of STEM graduates in the world (UNCTAD, 2018), the stream choice of students has not been sufficiently studied in the literature. In the Indian context, where specializing in science stream at the higher secondary is a mandatory prerequisite for STEM choices at the subsequent tertiary levels of education, understanding the stream choice at the higher secondary level becomes a policy imperative.

## **2.2 Caste and Education Outcomes in India**

Caste in India is a salient social identity. Caste significantly affects the opportunities of an individual's economic life: schooling, college, labour market, as well as old age (see Mosse, 2018; Munshi, 2019 for a review). The hierarchical nature of the caste system has generated historical social and economic inequalities between the caste groups. These inequalities have persisted in educational attainment between the caste groups, which further contribute to wage inequalities (Dutta & Das, 2007). Bailwal & Paul (2021) find that villages with a higher share of marginalized castes, i.e. SCs and STs have a lower probability of having public schools. Even in private schools, the disadvantaged caste groups have lower attendance which is significantly explained by the variations in parental education and household consumption expenditure (Azam, 2017). Although the policy of affirmative action has contributed to the reduction of gaps in intergenerational education mobility between the caste groups

(Hnatkovska et al., 2013), caste as a social identity and its role in establishing socio-cultural hierarchy is found to significantly affect not just the learning outcomes in school level children (Borooah, 2012; Hoff & Pandey, 2006) but well into the uppermost echelons of higher education (Subramanian, 2019).

While the role of castes in learning, educational attainment, and labour market outcomes has been explored in the literature, its role in STEM choice has not been explored. Much of the literature on STEM gaps being based on developed countries, racial segregation in STEM has been the focus of these studies (Xie et al., 2015). It is important to note that caste as a social identity is different from race. The caste system, being predominantly occupation-based, has been shown to restrict occupational mobility (Deshpande & Palshikar, 2008; Iversen et al., 2016). There are several channels through which caste can influence STEM choice other than historical inequities in wealth and education. Firstly, caste networks play a significant role in occupational choices. Munshi & Rosenzweig (2006) find that the institution of caste system shapes the career choices of men and women of the marginalized caste groups through the social networks of traditional occupations. Secondly, behavioral traits such as competitiveness, confidence, and grit, which have been shown to increase the likelihood of STEM participation are found to be lower among the historically marginalized caste groups: SC, ST, and OBC groups compared to the OC group (Banerjee et al., 2018; Dasgupta et al., 2020). Thirdly, the disadvantaged caste groups have been found to have lower levels of aspirations, perceived income adequacy, and actual earnings (Goel & Deshpande, 2020; Sarkar et al., 2020). Lower income aspirations could potentially shape the choices of STEM education and careers, which have been found to have higher returns in the labour market. While the caste identity could shape STEM choices through these mechanisms, the caste-based gaps in STEM fields of study have not been largely explored, which this paper seeks to address.

### **2.3 Gender and Education Outcomes in India**

While there has been significant progress in the reduction of gender gaps in education participation worldwide, gender differences in STEM education have persisted at all levels of education (UNESCO, 2017). The role of gender identity in STEM choice has been widely researched (Kahn & Ginther, 2017). Beginning at the school level, significant gender gaps in math and science achievements, which are considered as necessary skills for a STEM career, have been documented (Fryer & Levitt, 2010; Mullis et al., 2000). However, there is growing consensus that these differences in math and science performance are insignificant and not



reflective of gender differences in math or science ability (Lindberg et al., 2010). Studies have shown that the differences in test performance are often driven by gender stereotypes held by teachers (Carlana, 2019; Lavy & Sand, 2018; Rakshit & Sahoo, 2021) and by parents (Nollenberger et al., 2016). At the university level, gender differences in STEM preference and interests have been shown to explain the gender gaps in STEM-related major choices (Wiswall & Zafar, 2018; Zafar, 2013). The “chilly climate” effect experienced by women in male-dominated STEM occupations induces a lower preference for such occupations for women (Lordan & Pischke, 2016) and also explains the higher exit rates of women from STEM (Hunt, 2016).

In the Indian context, it is well documented that the gender gap in educational attainment is driven by the intra-household differences in the human capital investments based on the gender of the child (Kingdon, 2002; Kugler & Kumar, 2017; Rammohan & Vu, 2018; Sahoo, 2017). Although the under-representation of women in science has been widely studied globally, studies focused on the gender gaps in STEM education in the Indian context are few, especially at the higher secondary level. Sahoo & Klasen (2021) find that girls are less likely than boys to study science or commerce even after controlling for measures for cognitive ability. They also find that better access to STEM-related education benefits girls more than boys. In this paper, we use a larger sample from the nationally representative NSS surveys of three rounds spanning 2007-08 to 2017-18 to document and analyze the interaction of gender and caste in the context of STEM choice.

## **2.4 Interaction of Multiple Identities**

While considering the role of social identities in STEM choice it is important to consider the interaction of multiple identities for individuals who find themselves at the intersection of these identities, such as girls belonging to the disadvantaged caste groups. From the literature focused on wage discrimination for black women who suffer both racial and gender discrimination, the initial approach was to consider racial inequality and gender inequality independently. This approach had an implicit assumption that racial disadvantages were the same for both men and women, and the gender disadvantage faced by women was the same for each of the racial groups. Thus, black women faced the “double jeopardy” of race and gender, where the disadvantages due to each identity add up (Beale, 1979; King, 1988). This implied that the wage penalty faced by black women would be the summation of the race penalty (of being black) and gender penalty (of being a woman). While the “double jeopardy” framework has

been useful to explain the lowest wages of black women (Browne, 2000) and their low positions in organizational hierarchies (Maume, 1999), there have been concerns about the “additivity” of disadvantages as it ignores the interaction of multiple identities and treats each identity as separate and independent, having no complementarities (Bowleg, 2008; Browne & Misra, 2003).

With the emphasis on the complexities of intersecting identities, it is important to treat each intersecting group as unique (Brewer et al., 2002; McCall, 2005). Thus, an alternative approach in the literature has been the “intersectionality” analysis, which compares each intersecting group separately to a reference group (Greenman & Xie, 2008). This approach allows for disadvantages between two groups viz. the gender discrimination to vary across racial groups and vice versa. With respect to earnings in the US context, the gender gap in earnings is significantly lower for blacks and other ethnic minority groups compared to whites (Greenman & Xie, 2008).

## **2.5 Caste and Gender Intersectionality in India**

In the Indian context, the interaction of caste and gender identities becomes salient for the disadvantages faced by the women of marginalized castes. While women, in general, have lower educational attainment and labour force participation, there is substantial variation between women of different caste groups with SC and ST women experiencing most deprivation (Deshpande, 2007). In measures of sex ratio and labour force participation, ST women face lower discrimination than women from SC and other groups (Mitra, 2008). Women’s share of market work in a household has been found to decline as the social status of the caste group increases, suggesting that emphasis on family status amongst the less-disadvantaged castes detrimentally affects women’s participation in market work (Eswaran et al., 2013). Studies have also highlighted the difference between ST and SC women, with ST women having higher social status within their families considering labour market participation, age at marriage, mobility, and agency in intrahousehold decisions (Maity, 2017). The benefits of affirmative action have improved the educational attainments of the marginalized caste groups, but these gains have been concentrated among males with lesser effect for females (Cassan, 2019). These studies indicate how gender relations are affected by caste identities and the gender gaps vary across the caste groups, indicating the relevance of the intersectionality framework. However, analyzing the differences in aspirations and expectations of future wages for the participants of a skill training program, Sarkar et al. (2020)

find strong evidence of additive disadvantages of being female and belonging to the marginalized caste groups of SC/ST, lending support to the double jeopardy framework. We further contribute to this literature on intersectionality, focusing on the interaction of caste and gender in the context of STEM choice.

### **3. Data Description and Summary Statistics**

We use the data from three rounds of National Sample Survey Office (NSSO) led surveys on participation and expenditure on education.<sup>6</sup> These surveys collected relevant household-level information such as social groups, religion, usual consumer expenditure, place of residence, etc. Within households, the data also have information on individuals' gender, age, and details of educational enrollment. For the enrolled individuals, further information on the course enrolled, subject/stream chosen, and various educational expenses is available. The variable of interest for our study is the stream choice (science) of the students enrolled at the higher secondary level of schooling.

In India, children begin their primary schooling at the age of 6 years. Considering that students enroll in higher secondary after 10 years of primary and secondary schooling, we expect the age group of 16-18 years to be studying at the higher secondary level. The higher secondary enrollment rates among the children aged 16-18 years have been 22.2, 34.4, and 40.5 percent in 2007-08, 2014, and 2017-18, respectively.<sup>7</sup> While early enrollment in primary school at age of 5 years is possible in some states, late enrollment in primary school is common. We, therefore, consider a wider age band of 15-20 years of the students enrolled in the higher secondary level for analysis of stream choice. In the data pooled across the three rounds, we find that 96.53 percent of those enrolled in higher secondary level fall in this age category of 15-20 years.<sup>8</sup>

---

<sup>6</sup> The details of the survey samples are provided in the appendix (Table A1).

<sup>7</sup> Enrollment rates by gender and caste for this age group are provided in Table A2 in the appendix. While there is almost no gender gap in higher secondary enrollment rates, caste gaps are visible in every round. In a robustness analysis, we show that our regression analysis identifying the gender and caste gaps in science choice is not biased by any sample selection issue arising from non-random enrollment in higher secondary level.

<sup>8</sup> We also do our analysis on the sample restricted to those in the age group of 16-18 years and find that the results are almost identical, which is not surprising considering that 82.28 percent of our sample is from the age group of 16-18 years. These results are available on request.

The final sample for stream choice analysis considers the individuals enrolled in higher secondary level and within the age band of 15 and 20 years. Thus, we consider 6161 students for 2007-08 (64<sup>th</sup> round), 9076 students for 2014 (71<sup>st</sup> round), and 14697 students from 2017-18 (75<sup>th</sup> round), who were enrolled in the higher secondary level for our analysis. Pooling across the years, we have a sample for 29,934 students. For this sample, we construct a binary variable indicating enrollment in the science stream versus enrollment in the humanities or commerce stream.

Table 1 provides the summary statistics of the variables used in our analysis. We find that the proportion of students taking up science increased from 30 percent in 2007-08 (64<sup>th</sup> round) to 39.6 percent in 2014 (71<sup>st</sup> round). In 2017-18 this proportion is 38 percent, which is, although marginally lower, statistically not distinguishable from the proportion in 2014. While we observe that science take-up has substantially increased since 2007, there have been persistent gaps between male and female students as well as between the students belonging to different caste groups. Figure 1 shows a gender gap of about 7-10 percentage points in science take-up rates persistent over the years. In Figure 2, the proportion of students choosing science is presented for different caste groups. The science take-up by students of the SC group is persistently 6-7 percentage points lower compared to the OC group. The gap is even larger for the students of the ST group where the gap ranges between 16 and 20 percentage points. However, OBC students are found to be as likely as OC students to study science. To understand whether these gaps are driven by differences in observable characteristics across the groups, we further analyze the nature of these gaps in the subsequent sections.

#### **4. Methodology**

We conduct three types of empirical analyses. First, we quantify the gaps in science choice between gender- and caste-based groups using a linear probability model. Second, we explore these gaps using the Blinder-Oaxaca decomposition method (Blinder, 1973; Oaxaca, 1973). The decomposition methods are used to analyze the differences in a distributional statistic in terms of its explanatory factors (see Fortin et al., (2011) for an overview). Third, we investigate the intersectionality between caste and gender and test the double jeopardy hypothesis to document the additive effects of disadvantages faced by individuals in science choice.

## 4.1 Regression Analysis

We use regression analysis to estimate the gaps in science choice between the groups controlling for individual and household characteristics along with regional fixed effects. Pooling the three rounds of repeated cross-sectional data, we estimate the probability of taking up science stream for the sample of students enrolled in higher secondary level using a linear probability model.<sup>9</sup>

$$Sci_{idt} = \alpha_0 + \alpha_1 Fem_{idt} + \sum_j \beta_j Caste_{ijdt} + \gamma X_{idt} + \mu_d + \lambda_t + \varepsilon_{idt} \quad (1)$$

where  $Sci_{idt}$  is an indicator variable for science choice which takes the value of 1 if the student  $i$  in district  $d$  and time (survey-round)  $t$  has opted for science stream, and takes the value of 0 if the student has opted for other streams such as commerce or humanities.  $Fem_{idt}$  is an indicator variable for the student being female.  $Caste_{ijdt}$  takes value of 1 if the student  $i$  belongs to caste group  $j$ , where  $j \in \{SC, ST, OBC\}$ , the reference group being the students belonging to the OC.  $X_{idt}$  is a vector of student characteristics such as age, religion, household size, years of schooling of the head of household, monthly per capita consumption expenditure, and the household location being in the urban or rural region.  $\mu_d$  are district fixed effects,  $\lambda_t$  are the survey-round fixed effects and  $\varepsilon_{idt}$  is the error term. The use of district-level fixed effects is to control for observable and unobservable district-level factors, e.g. the demand for education and the availability of schools offering various streams, that could influence the stream choice at the higher secondary level. The survey-round fixed effects control for any secular change in stream choice over the years. We estimate equation (1) with the pooled cross-sectional data from three rounds.

## 4.2 Decomposition Analysis

To further analyze the gaps in science choice between caste groups and gender groups, we use the Blinder-Oaxaca technique that decomposes the differences in the science take-up rates between the groups into differences in observable characteristics (i.e. the differences in the endowment of explanatory factors, called the “endowment effect”) and into differences due to

---

<sup>9</sup> We present linear probability model since we are interested in the marginal effects that are straightforward to interpret from linear models. However, we also check the results using nonlinear (probit and logit) models and find similar results (presented in the appendix). Another concern is that this sample is based on students who are enrolled at the higher secondary level, hence, there may be sample selection bias. In a robustness analysis presented later, we estimate a sample selection model to consider the enrollment decision. The main results are unchanged.

the effect of those characteristics (i.e. the differences in their coefficients, called the “coefficient effect”).

$$Gap = \overline{Sci_A} - \overline{Sci_B} = (\overline{X_A} - \overline{X_B})\beta_B + \overline{X_A}(\beta_A - \beta_B) \quad (2)$$

The first term of (2) is called the endowment effect, and the second term is called the “coefficient effect”. This decomposition (2) is based on the viewpoint of group B. Conversely, this decomposition can also be re-arranged from the viewpoint of group A with different results. This is known in the decomposition literature as the “index number problem”, where the choice of the reference group affects the decomposition results. An alternative decomposition method that is prominent in the literature is using a non-discriminatory coefficient  $\beta^*$ ,

$$Gap = \overline{Sci_A} - \overline{Sci_B} = (\overline{X_A} - \overline{X_B})\beta^* + [\overline{X_A}(\beta_A - \beta^*) + \overline{X_B}(\beta^* - \beta_B)] \quad (3)$$

The first term of (3) is part of the gap that is “explained” by the differences in the explanatory variables. The summation of the second and third terms of (3) is called the “unexplained” part of the gap. We follow Neumark (1988) in our choice of non-discriminatory  $\beta^*$ , which is estimated from the pooled regression including both groups A and B. The unexplained term in the decomposition is calculated by subtracting the “explained” part from the overall gap.

We estimate the decomposition (3) for both the gender groups and caste groups.<sup>10</sup> For gender, the reference group is the male students (group A) and we decompose the gaps observed between the male students and the female students (group B) in science choice. For the caste groups, the reference group is the students of OC (group A) and we decompose the observed differences in science choice between the students of OC and each of the other caste groups (group B): OBC, SC, and ST.

### 4.3 Intersectionality Analysis

Next, we analyze the interaction of caste and gender identities in determining the science choice of female students from the marginalized caste groups. To test the double jeopardy versus intersectionality hypothesis, we include dummy variables denoting whether an individual belongs to a group defined by the intersection of the two identities. Considering the two gender-

---

<sup>10</sup> In the decomposition analysis, since our interest is to analyze how much of the gaps are explained by endowments, we use a specification that controls for observable individual and household specific characteristics along with survey-round fixed effects.

based groups (male and female) and four caste-based groups (SC, ST, OBC, and OC), eight such groups lie at the intersection. With the comparison group being the male students from the OC group, the augmented regression model now includes seven dummy variables to define the identity of individuals: three for males belonging to OBC, SC, and ST groups, and four for females from the OC, OBC, SC, and ST groups.

$$\begin{aligned} Sci_{idt} = & \alpha_0 + \alpha_1 Male\_OBC_{idt} + \alpha_2 Male\_SC_{idt} + \alpha_3 Male\_ST_{idt} \\ & + \beta_0 Fem\_OC_{idt} + \beta_1 Fem\_OBC_{idt} + \beta_2 Fem\_SC_{idt} + \beta_3 Fem\_ST_{idt} \\ & + \gamma X_{idt} + \mu_d + \lambda_t + \varepsilon_{idt} \end{aligned} \quad (4)$$

The hypothesis of double jeopardy would be supported if the coefficients  $\alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2$  and  $\beta_3$  are negative and significant, with  $\beta_j \approx \beta_0 + \alpha_j$ , for  $j \in \{1,2,3\}$ . For example, in comparison with the dominant group, i.e. male students of the OC group, the gap in science choice for female students of SC groups ( $\beta_2$ ) would be approximately equal to the summation of gaps seen in female students of the OC group ( $\beta_0$ ) and males of the SC group ( $\alpha_2$ ). This would indicate that the females of the SC group are faced with the double disadvantage of being from the SC group and being female.

Alternatively, the disadvantages of caste (as experienced by males of OBC, SC, and ST) captured by  $\alpha_j$  and the disadvantage of gender (as experienced by females of OC) captured by  $\beta_0$  may not exactly add up for individuals lying at the intersection of these identities. In such a case, where  $\beta_j \neq \beta_0 + \alpha_j$ , for  $j \in \{1,2,3\}$ , the disadvantage faced by the females of OBC, SC, and ST groups would be different from the summation of the caste and gender disadvantages. This signifies that the experience of the groups at the intersection of the two identities is unique and significantly different from what is captured only by adding the effects of two identities. This is known as intersectionality in the literature. We can also express the intersectionality, i.e. the effect of the intersection of two identities as  $\delta_j = \beta_j - (\alpha_j + \beta_0) \neq 0, j \in \{1,2,3\}$ . An estimate of  $\delta_j$  being significantly different from zero would suggest evidence for the intersectionality effect.<sup>11</sup>

---

<sup>11</sup> Equivalently,  $\delta_j$  can also be estimated with the interaction of caste and gender dummies as in the following model:

$$Sci_{idt} = \alpha_0 + \alpha_1 Fem_{idt} + \sum_j \beta_j Caste_{ijdt} + \sum_j \delta_j Caste_{ijdt} * Fem_{ijdt} + \gamma X_{idt} + \mu_d + \lambda_t + \varepsilon_{idt}$$

## 5. Results

### 5.1 Regression Analysis

We report the results of estimating equation (1) in Table 2.<sup>12</sup> Column 1 presents the estimates from the parsimonious specification that includes only the survey-round fixed effects. Subsequently, we add individual and household characteristics along with state fixed effects (column 2) and district fixed effects (column 3). We further include household fixed effects in the final specification to identify the intra-household gender gap, although this model does not allow us to identify the effect of any household-specific factor including caste, as they are collinear with the fixed effects. We also perform the regression analysis separately for each round (as per the specification in column 3 of Table 2) and plot the coefficients for female and caste-group dummies in Figure 3.<sup>13</sup>

For the pooled data, across all specifications, we find a significant negative coefficient for female students. The inclusion of control variables in columns 2 and 3 does not cause much variation in either the magnitude or the statistical significance of the gender gap. We find that girls are around 9 percentage points less likely than boys to choose the science stream. The magnitude of the gender gap is quite substantial as it implies a 25 percent gap with respect to the mean participation rate in science (which is 36.7 percent among all the students). Further, in column 4, when household fixed effects are included in the specification, we find a significant gender gap of 9.4 percentage points.<sup>14</sup> This evidence suggests that the gender gap in science choice is also an intra-household phenomenon. From the round-wise regressions, we find that the gender gap is significant and between 8 and 11 percentage points for each round (Figure 3). Thus, the gender gap has persisted with no substantial change over the years.

Focusing on the caste gaps, we find that students from SC, ST, and OBC groups are significantly less likely to choose science than students from the other caste (OC) group. In terms of hierarchical distance in the traditional ordering of caste groups, the OBC is the closest to the OC. While the raw differences in science choice are not significant between OC and

---

<sup>12</sup> We replicate the estimation using non-linear models (probit and logit) and find similar results both in magnitude and the significance of the gaps. The marginal effects from the non-linear estimations are provided in Table A7 in the appendix.

<sup>13</sup> We also report the full results of round-wise estimation in tables A3–A5 in the appendix.

<sup>14</sup> Inclusion of household fixed effects implies that the estimation sample considers only those households where there are at least two individuals enrolled in higher secondary level. Therefore, the sample size reduces.



OBC, once we control for observable factors and regional fixed effects, we find a significant negative effect for the OBC group in science choice. The results suggest that the OBC students are about 2.7 percentage points less likely to pursue science courses in high school, controlling for observables and district fixed effects. This gap is even more stark for the students of SC and ST groups, as they are 4.4 and 6.9 percentage points, respectively, less likely to pursue science than those of the OC group. To reflect on the size of these effects, they imply a gap of 12 percent for SC and 19 percent for ST students, with respect to the mean participation rate in the science stream. For these two social groups, we also notice that the magnitude of the coefficient declines as we include more control variables in the regression, suggesting that the control variables capture some of the socio-economic disadvantages that drive the lower probability of science choice for these groups. We delve deeper into this issue using a decomposition analysis presented in the subsequent section.

There may be significant differences in gender norms and social discrimination between rural and urban areas. Hence, we also estimate all four specifications of Table 2 separately for the rural and urban samples. The results are presented in Table A6 in the appendix. As gender norms and caste hierarchies are stronger in the rural areas, we expect the gaps to be larger in the rural sample. Indeed, we find that both the gender and caste gaps are larger and more significant in the rural regions. While the gender gap in urban areas is nearly half of that in rural areas, the caste gaps in urban areas are insignificant.

## **5.2 Decomposition Analysis**

In this section, we further investigate the nature of gender and caste gaps using decomposition analysis. The results of decomposition are presented in Table 3. Column (1) presents the results for the gender gap where the overall gap of 9 percentage points is decomposed into a part that is explained by the differences in observable characteristics and the remaining part that is not explained by them. The explained part is found to be negligible in magnitude and statistically not significant. This result suggests that the gender gap in science choice is not driven by the differences in endowments, but by other unobservable factors.

Next, we decompose the caste gap between each of the disadvantaged groups (SC, ST, and OBC) and the non-disadvantaged group (OC). Unlike the gender gap, here we find that the differences in endowments significantly explain some parts of the gaps. Column (2) presents the results of decomposition of the gap between the OC and OBC groups. OBC group on average has lower endowments in terms of household head's education and per capita

consumption expenditure, which contribute to the explained part of the gap being positive and significant. Columns (3) and (4) report the results for SC and ST groups respectively. Both the SC and ST groups have significantly lower science take-up rates compared to the OC group, and a substantial part of these gaps is explained by the differences in endowments.

We further analyze the contribution of the control variables in the explained part of the caste gaps. We find that two variables – education (years of schooling) of the head of the household and monthly per capita expenditure (MPCE) – substantially contribute to the explained differences. Thus, the caste gap in science choice is explained by the fact that SC/ST students come from households with lower economic affluence and where the household head has lower educational attainment compared to the non-disadvantaged social groups (OC). Thus, the historical disadvantages in schooling for earlier generations and lower incomes of the disadvantaged caste groups partly drive the gaps in science choice.

Unlike the caste-based gaps, the gender gap in science choice is not significantly explained by the differences in endowments. Considering the two variables that were explaining the caste-based gaps, we find that in the case of the gender gaps, the education of the head of the household is not significant. Although the explained difference by MPCE is significant, it is small in magnitude (only about 3 percent of the gender gap). With the overall explained differences for gender gaps being not significant, the gap could be driven by unobservable factors such as preferences and gender norms, which we cannot test given the limitation of our data. Overall, our results suggest that the underlying factors driving the gender gap are different from those driving the caste gaps.

### **5.3 Intersectionality Analysis**

In this section, we analyze the interaction of the gaps in science choice due to caste and gender identities. We test for the double jeopardy hypothesis using model (4) and report the results in Table 4. Column (1) reports the regression results without any control variable except survey-round dummies. Column (2) includes the control variables and state fixed effects, and column (3) includes district fixed effects. In panel A, we report the coefficient for each of the intersecting gender and caste groups. For evidence supporting double jeopardy, the coefficients for females from OBC, SC, or ST groups should be approximately equal in magnitude to the summation of the coefficient for females from OC group and the coefficient for males from their respective caste groups (i.e. OBC, SC, or ST groups, respectively). On the other hand, if they are not equal in magnitude, then the result would support the presence of intersectionality.

Hence, we report these linear combinations of the coefficients and test their statistical significance in panel B.

From the results in panel B, we are unable to reject the double jeopardy hypothesis for female students of the OBC and SC groups. This lends credence to the cumulative disadvantages faced by the females in these caste groups. However, for the females of the ST group, we find that the linear combination of coefficients is around 0.07 that is significantly different from zero (columns 2 and 3 of panel B). In other words, in terms of absolute magnitude, the coefficient of female-ST students is significantly lower than the summation of the coefficients of male-ST and female-OC students. This result implies that in the case of females belonging to the ST group, we can reject the double jeopardy hypothesis in favour of intersectionality.<sup>15</sup> This result also implies that the gender gap among the ST students is lower in magnitude than the gender gap among the OC students. This finding is consistent with the literature highlighting lower gender discrimination among the ST group (Maity, 2017; Mitra, 2008). However, in our context, the smaller gender gap among ST students may also be due to the low enrollment rates in the science stream among both male and female students of the ST group.

#### **5.4 Robustness Check – Sample Selection Correction**

Our empirical model (1) is estimated on the sample of students currently enrolled in the higher secondary level of education. Stream choice is not observed for individuals who are not enrolled in higher secondary schooling. This may raise concerns about sample selection bias, especially if unobservable factors driving the decision of enrollment are also correlated with science choice (Heckman, 1979). To address this issue, we estimate a sample selection model that considers both enrollment and stream choice decisions.

The selection model is estimated in two steps. First, a probit model is estimated considering the binary choice of enrolling into higher secondary education.<sup>16</sup> From this model, a selection

---

<sup>15</sup> We also estimate the intersectionality hypothesis using interaction terms of female and caste dummies, and report the results in Table A8 in the appendix. The results are equivalent to what we present in Table 4.

<sup>16</sup> We consider all individuals in the 15-20 age-group for the enrollment decision. The household fixed effects specification may suffer from the incidental parameters problem due to the inclusion of a large number of fixed effects in a probit model. Hence, we follow Wooldridge (1995) and include the Mundlak terms (i.e. household level averages of the individual level variables) in this model. We also interact the explanatory variables with year fixed effects in the probit model. This accounts for the fact that enrollment rates have increased over the years and hence, the effect of the explanatory factors on enrollment may also vary over time.

correction term, known as the Inverse Mills Ratio (IMR), is calculated. In the second step, taking the sample of enrolled students, we estimate a linear probability model for science choice where the IMR is included as an additional explanatory variable. Due to the inclusion of IMR which is a generated regressor, we use non-parametric bootstrapping to obtain the standard errors for the science choice equation.<sup>17</sup>

We report the selection-corrected estimates in Table 5. Consistent with the results in Table 1, we find that the gender gap varies between 9-11 percentage points and is statistically significant across specifications. The caste gaps in science choice are also statistically significant and very close to the effect sizes estimated in our baseline model. The selection correction term is statistically significant, especially for the year 2007-08 when the enrollment rate in higher secondary level was relatively lower. However, our main findings on the gender and caste gaps in science choice remain robust.

## 6. Conclusion

Our study quantitatively documents the role of social identities of gender and caste in STEM choice at the higher secondary level of education in India. We show that these social identities significantly influence the science stream choice even after controlling for various observable factors at the household level and regional fixed effects. While the effect of caste-based identities can be in part explained by differences in household resources between the caste-based social groups, the effect of gender identity remains largely unexplained by the differences in observable characteristics. While social identities could affect STEM choices through mechanisms such as stereotype threats, social networks, behavioral traits, or aspirations, the limitation of our data does not allow us to test these channels. We also show that the interaction between gender and caste identities results in additive disadvantages for the female students of the SC and the OBC groups but not for the ST group. Thus, the social identity of gender affects choices differently from the social identity of caste, and the

---

<sup>17</sup> For ease of identification, the probit model for enrollment should have an explanatory variable that is validly excluded from the science choice equation. For this purpose, we consider a binary variable indicating whether a household fell in the bottom quartile of the MPCE distribution. In an era when education enrollment has expanded substantially, only those belonging to the poorest households are likely to not enroll. Indeed, this indicator variable shows a significant and negative effect on enrollment. The linear MPCE variable is included in both enrollment and stream choice equations. We also interact IMRs with year dummies in the science choice equation to allow the selection effect to vary across the years, which is important considering the steady rise in enrollment over time.

interaction of these two identities results in unique disadvantages for some groups at the interaction of these identities.

Ensuring equality in access to affordable and quality technical education is one of the targets of the Sustainable Development Goals for 2030. In the context of a large developing country, we show how gender and caste-based identity adversely affects participation in STEM education that has significant labour market returns, and in turn, has implications for economic welfare and social mobility. Thus, our study underscores the importance of considering social identity in devising policies to promote equality in educational opportunities.

## References

- Agrawal, T. (2014). Gender and caste-based wage discrimination in India: Some recent evidence. *Journal for Labour Market Research*, 47(4), 329–340.  
<https://doi.org/10.1007/s12651-013-0152-z>
- Altonji, J. G., Arcidiacono, P., & Maurel, A. (2016). Chapter 7 - The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 5, pp. 305–396). Elsevier. <https://doi.org/10.1016/B978-0-444-63459-7.00007-5>
- Azam, M. (2017). Explaining Caste Differences in Private School Attendance. *Review of Development Economics*, 21(4), 1191–1204. <https://doi.org/10.1111/rode.12305>
- Bailwal, N., & Paul, S. (2021). Caste Discrimination in Provision of Public Schools in Rural India. *The Journal of Development Studies*, 0(0), 1–22.  
<https://doi.org/10.1080/00220388.2020.1862796>
- Banerjee, R., Gupta, N. D., & Villeval, M. C. (2018). The spillover effects of affirmative action on competitiveness and unethical behavior. *European Economic Review*, 101, 567–604. <https://doi.org/10.1016/j.eurocorev.2017.10.022>
- Beale, F. (1979). Double Jeopardy: To Be Black and Female in The Black Woman: An Anthology (ed.) Toni Cadre. *New York: New American Library*.
- Beffy, M., Fougère, D., & Maurel, A. (2012). CHOOSING THE FIELD OF STUDY IN POSTSECONDARY EDUCATION: DO EXPECTED EARNINGS MATTER? *The Review of Economics and Statistics*, 94(1), 334–347.

- Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4), 436–455. <https://doi.org/10.2307/144855>
- Borooah, V. K. (2012). Social Identity and Educational Attainment: The Role of Caste and Religion in Explaining Differences between Children in India. *The Journal of Development Studies*, 48(7), 887–903. <https://doi.org/10.1080/00220388.2011.621945>
- Boudarbat, B., & Montmarquette, C. (2009). Choice of fields of study of university Canadian graduates: The role of gender and their parents' education. *Education Economics*, 17(2), 185–213. <https://doi.org/10.1080/09645290802133032>
- Bowleg, L. (2008). When Black+ lesbian+ woman ≠ Black lesbian woman: The methodological challenges of qualitative and quantitative intersectionality research. *Sex Roles*, 59(5), 312–325.
- Brewer, R. M., Conrad, C. A., & King, M. C. (2002). The Complexities and Potential of Theorizing Gender, Caste, Race, and Class. *Feminist Economics*, 8(2), 3–17. <https://doi.org/10.1080/1354570022000019038>
- Browne, I. (2000). *Latinas and African American women at work: Race, gender, and economic inequality*. Russell Sage Foundation.
- Browne, I., & Misra, J. (2003). The Intersection of Gender and Race in the Labor Market. *Annual Review of Sociology*, 29(1), 487–513. <https://doi.org/10.1146/annurev.soc.29.010202.100016>
- Carlana, M. (2019). Implicit Stereotypes: Evidence from Teachers' Gender Bias\*. *The Quarterly Journal of Economics*, 134(3), 1163–1224. <https://doi.org/10.1093/qje/qjz008>
- Cassan, G. (2019). Affirmative action, education and gender: Evidence from India. *Journal of Development Economics*, 136, 51–70. <https://doi.org/10.1016/j.jdeveco.2018.10.001>
- Crenshaw, K. (1990). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, 43(6), 1241–1300.
- Dasgupta, U., Mani, S., Sharma, S., & Singhal, S. (2020). *Social Identity, Behavior, and Personality: Evidence from India* (SSRN Scholarly Paper ID 3660261). Social Science Research Network. <https://papers.ssrn.com/abstract=3660261>
- Deshpande, A. (2007). Overlapping Identities under Liberalization: Gender and Caste in India. *Economic Development and Cultural Change*, 55(4), 735–760. <https://doi.org/10.1086/516763>
- Deshpande, A. (2011). *The Grammar of Caste: Economic Discrimination in Contemporary India*. Oxford University Press.

- Deshpande, R., & Palshikar, S. (2008). Occupational Mobility: How Much Does Caste Matter? *Economic and Political Weekly*, 43(34), 61–70.
- Dutta, P. V., & Das, M. B. (2007). *Does caste matter for wages in the Indian labor market?* The World Bank.
- Eswaran, M., Ramaswami, B., & Wadhwa, W. (2013). Status, Caste, and the Time Allocation of Women in Rural India. *Economic Development and Cultural Change*, 61(2), 311–333. <https://doi.org/10.1086/668282>
- Fortin, N., Lemieux, T., & Firpo, S. (2011). Chapter 1—Decomposition Methods in Economics. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4, pp. 1–102). Elsevier. [https://doi.org/10.1016/S0169-7218\(11\)00407-2](https://doi.org/10.1016/S0169-7218(11)00407-2)
- Fryer, R. G., & Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, 2(2), 210–240. <https://doi.org/10.1257/app.2.2.210>
- Gemici, A., & Wiswall, M. (2014). Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors. *International Economic Review*, 55(1), 23–56. <https://doi.org/10.1111/iere.12040>
- Goel, D., & Deshpande, A. (2020). Social identity and perceived income adequacy. *Review of Development Economics*, 24(2), 339–361. <https://doi.org/10.1111/rode.12646>
- Greenman, E., & Xie, Y. (2008). Double Jeopardy? The Interaction of Gender and Race on Earnings in the United States. *Social Forces*, 86(3), 1217–1244. <https://doi.org/10.1353/sof.0.0008>
- Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153–161. <https://doi.org/10.2307/1912352>
- Hnatkovska, V., Lahiri, A., & Paul, S. B. (2013). Breaking the Caste Barrier Intergenerational Mobility in India. *Journal of Human Resources*, 48(2), 435–473. <https://doi.org/10.3368/jhr.48.2.435>
- Hoff, K., & Pandey, P. (2006). Discrimination, Social Identity, and Durable Inequalities. *American Economic Review*, 96(2), 206–211. <https://doi.org/10.1257/000282806777212611>
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46), 16385–16389.
- Hunt, J. (2016). Why do Women Leave Science and Engineering? *ILR Review*, 69(1), 199–226. <https://doi.org/10.1177/0019793915594597>

- India, Ministry of Information and Broadcasting, Publications Division, India, Ministry of Finance, & Economic Division. (2021). *Economic survey 2020-21.: Vol. Volume 2, Chapter 9*.
- Iversen, V., Krishna, A., & Sen, K. (2016). Rags to riches? Intergenerational occupational mobility in India. In *Global Development Institute Working Paper Series* (No. 042016; Global Development Institute Working Paper Series). GDI, The University of Manchester. <https://ideas.repec.org/p/bwp/bwppap/042016.html>
- Jain, T., Mukhopadhyay, A., Prakash, N., & Rakesh, R. (2018). Labor Market Effects of High School Science Majors in a High STEM Economy. *IZA Discussion Paper No. 11934*, Institute of Labor Economics, Bonn, 44.
- Jensen, R. (2012). Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India \*. *The Quarterly Journal of Economics*, 127(2), 753–792. <https://doi.org/10.1093/qje/qjs002>
- Kahn, S., & Ginther, D. (2017). *Women and STEM* (No. w23525). National Bureau of Economic Research. <https://doi.org/10.3386/w23525>
- Kearney, C. (2011). *Efforts to Increase Students' Interest in Pursuing Science, Technology, Engineering and Mathematics Studies and Careers* (p. 48). European Schoolnet (EUN Partnership AISBL). [https://swsu.ru/sbornik-statey/pdf/2011\\_european\\_schoolnet.pdf](https://swsu.ru/sbornik-statey/pdf/2011_european_schoolnet.pdf)
- King, D. K. (1988). Multiple jeopardy, multiple consciousness: The context of a Black feminist ideology. *Signs: Journal of Women in Culture and Society*, 14(1), 42–72.
- Kingdon, G. G. (2002). The Gender Gap in Educational Attainment in India: How Much Can Be Explained? *The Journal of Development Studies*, 39(2), 25–53. <https://doi.org/10.1080/00220380412331322741>
- Kugler, A. D., & Kumar, S. (2017). Preference for Boys, Family Size, and Educational Attainment in India. *Demography*, 54(3), 835–859. <https://doi.org/10.1007/s13524-017-0575-1>
- Lavy, V., & Sand, E. (2018). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *Journal of Public Economics*, 167, 263–279. <https://doi.org/10.1016/j.jpubeco.2018.09.007>
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, 136(6), 1123–1135. <https://doi.org/10.1037/a0021276>

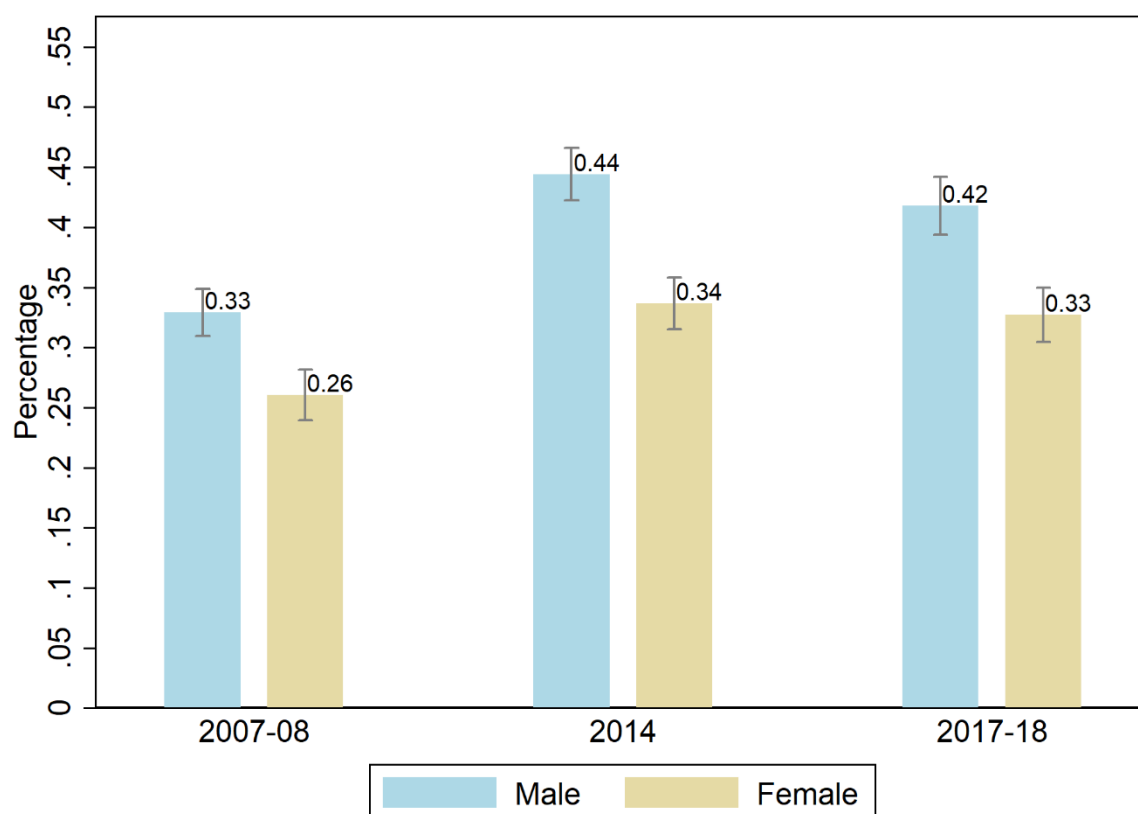


- Lordan, G., & Pischke, J.-S. (2016). *Does Rosie Like Riveting? Male and Female Occupational Choices* (Working Paper No. 22495; Working Paper Series). National Bureau of Economic Research. <https://doi.org/10.3386/w22495>
- Maity, B. (2017). Comparing Health Outcomes Across Scheduled Tribes and Castes in India. *World Development*, 96, 163–181. <https://doi.org/10.1016/j.worlddev.2017.03.005>
- Maltese, A. V., Melki, C. S., & Wiebke, H. L. (2014). The Nature of Experiences Responsible for the Generation and Maintenance of Interest in STEM. *Science Education*, 98(6), 937–962. <https://doi.org/10.1002/sce.21132>
- Maume, D. J. (1999). Glass Ceilings and Glass Escalators: Occupational Segregation and Race and Sex Differences in Managerial Promotions. *Work and Occupations*, 26(4), 483–509. <https://doi.org/10.1177/0730888499026004005>
- McCall, L. (2005). The Complexity of Intersectionality. *Signs: Journal of Women in Culture and Society*, 30(3), 1771–1800. <https://doi.org/10.1086/426800>
- Mitra, A. (2008). The status of women among the scheduled tribes in India. *The Journal of Socio-Economics*, 37(3), 1202–1217. <https://doi.org/10.1016/j.socec.2006.12.077>
- Mosse, D. (2018). Caste and development: Contemporary perspectives on a structure of discrimination and advantage. *World Development*, 110, 422–436. <https://doi.org/10.1016/j.worlddev.2018.06.003>
- Mullis, I. V. S., Third International Mathematics and Science Study, & International Association for the Evaluation of Educational Achievement (Eds.). (2000). *Gender differences in achievement: IEA's Third International Mathematics and Science Study (TIMSS)*. TIMSS International Study Center, Boston College.
- Munshi, K. (2019). Caste and the Indian Economy. *Journal of Economic Literature*, 57(4), 781–834. <https://doi.org/10.1257/jel.20171307>
- Munshi, K., & Rosenzweig, M. (2006). Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy. *American Economic Review*, 96(4), 1225–1252. <https://doi.org/10.1257/aer.96.4.1225>
- Neumark, D. (1988). Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. *The Journal of Human Resources*, 23(3), 279–295. <https://doi.org/10.2307/145830>
- Nollenberger, N., Rodríguez-Planas, N., & Sevilla, A. (2016). The Math Gender Gap: The Role of Culture. *American Economic Review*, 106(5), 257–261. <https://doi.org/10.1257/aer.p20161121>

- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693–709. <https://doi.org/10.2307/2525981>
- Oster, E., & Steinberg, B. M. (2013). Do IT service centers promote school enrollment? Evidence from India. *Journal of Development Economics*, 104, 123–135. <https://doi.org/10.1016/j.jdeveco.2013.05.006>
- Rakshit, S., & Sahoo, S. (2021). *Biased Teachers and Gender Gap in Learning Outcomes: Evidence from India*. IZA Discussion Paper No. 14305, Institute of Labor Economics, Bonn.
- Rammohan, A., & Vu, P. (2018). Gender Inequality in Education and Kinship Norms in India. *Feminist Economics*, 24(1), 142–167. <https://doi.org/10.1080/13545701.2017.1364399>
- Roemer, J. E., & Trannoy, A. (2015). Chapter 4—Equality of Opportunity. In A. B. Atkinson & F. Bourguignon (Eds.), *Handbook of Income Distribution* (Vol. 2, pp. 217–300). Elsevier. <https://doi.org/10.1016/B978-0-444-59428-0.00005-9>
- Roychowdhury, P. (2021). (Em)Powered by Science? Estimating the Relative Labor Market Returns to Majoring in Science in High School in India. *Economics of Education Review*, 82, 102118. <https://doi.org/10.1016/j.econedurev.2021.102118>
- Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education*, 96(3), 411–427. <https://doi.org/10.1002/sce.21007>
- Sahoo, S. (2017). Intra-Household Gender Disparity in School Choice: Evidence from Private Schooling in India. *The Journal of Development Studies*, 53(10), 1714–1730. <https://doi.org/10.1080/00220388.2016.1265943>
- Sahoo, S., & Klasen, S. (2018). *Gender Segregation in Education and Its Implications for Labour Market Outcomes: Evidence from India*. IZA Discussion Paper No. 11660, Institute of Labor Economics, Bonn.
- Sahoo, S., & Klasen, S. (2021). Gender Segregation in Education: Evidence From Higher Secondary Stream Choice in India. *Demography*, 9101042. <https://doi.org/10.1215/00703370-9101042>
- Sarkar, S., Chakravorty, B., & Lyonette, C. (2020). *Social Identity and Aspiration - Double Jeopardy or Intersectionality? Evidence from Rural India* (Working Paper No. 724). GLO Discussion Paper. <https://www.econstor.eu/handle/10419/226520>
- Shastri, G. K. (2012). Human Capital Response to Globalization: Education and Information Technology in India. *The Journal of Human Resources*, 47(2), 287–330.

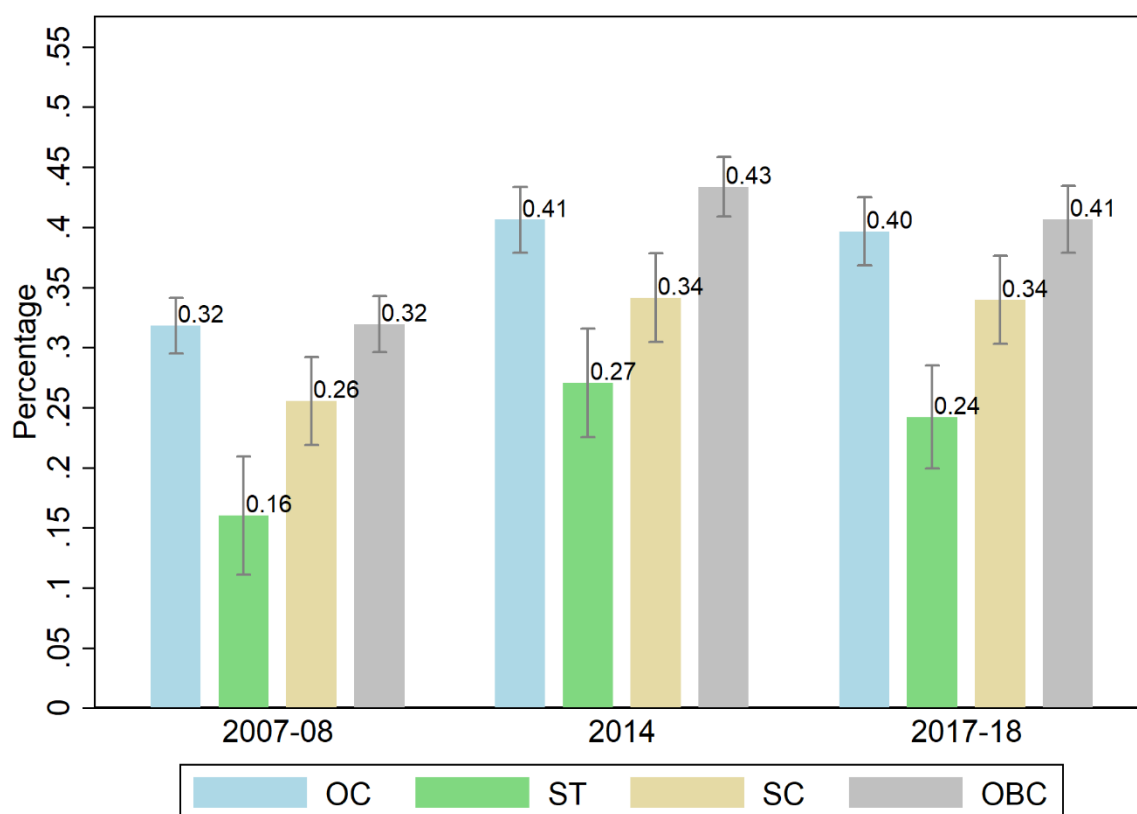
- Subramanian, A. (2019). *The Caste of Merit: Engineering Education in India*. Harvard University Press.
- UNCTAD (Ed.). (2018). *Harnessing frontier technologies for sustainable development*. UNCTAD.
- UNESCO. (2017). *Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)*. UNESCO.  
<https://unesdoc.unesco.org/ark:/48223/pf0000253479>
- White House Report. (2013). *Women and Girls in Science, Technology, Engineering, and Math (STEM)*. Executive Office of the President.  
[https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/women\\_and\\_girls\\_in\\_stem\\_factsheet\\_june\\_2013.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/women_and_girls_in_stem_factsheet_june_2013.pdf)
- Wiswall, M., & Zafar, B. (2018). Preference for the Workplace, Investment in Human Capital, and Gender\*. *The Quarterly Journal of Economics*, 133(1), 457–507.  
<https://doi.org/10.1093/qje/qjx035>
- Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68(1), 115–132.  
[https://doi.org/10.1016/0304-4076\(94\)01645-G](https://doi.org/10.1016/0304-4076(94)01645-G)
- Xie, Y., Fang, M., & Shauman, K. (2015). STEM Education. *Annual Review of Sociology*, 41(1), 331–357. <https://doi.org/10.1146/annurev-soc-071312-145659>
- Zafar, B. (2013). College Major Choice and the Gender Gap. *The Journal of Human Resources*, 48(3), 545–595. JSTOR.

Figure 1. Trends in science choice at higher secondary level by gender



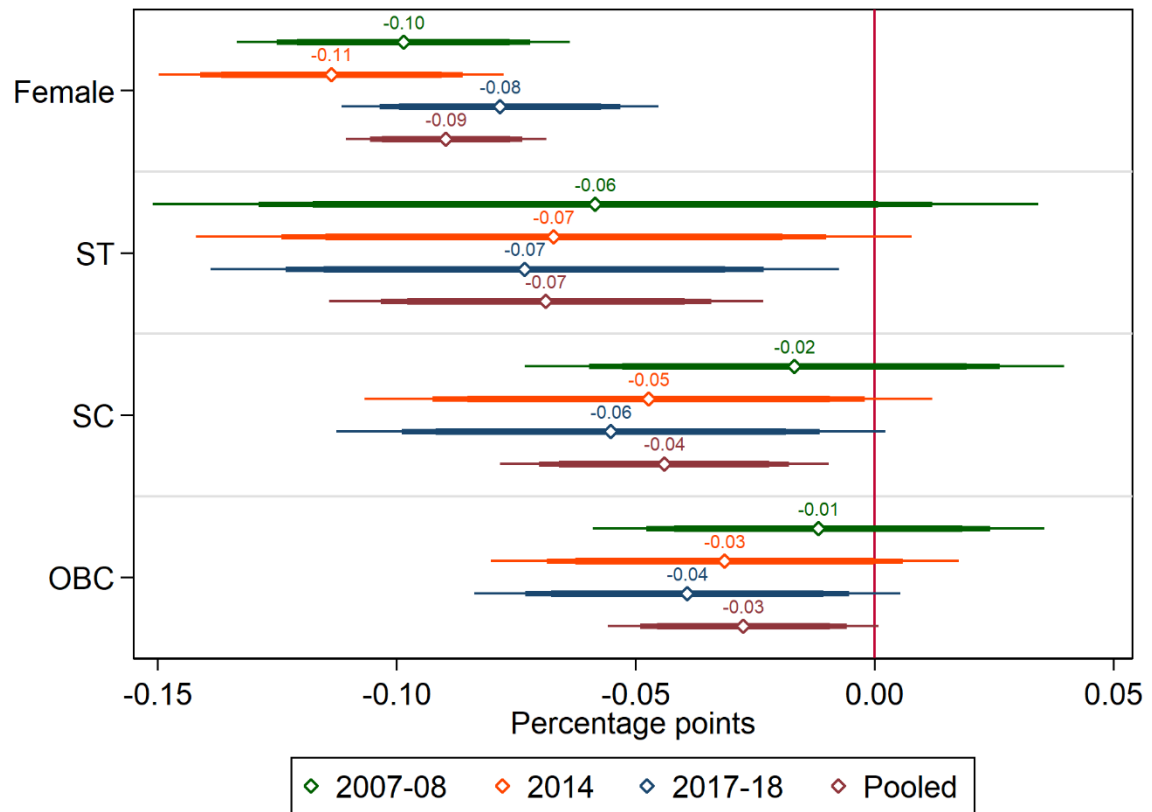
Note: The figures are estimated using data from multiple rounds of National Sample Survey data on education. Gender wise, the percentage of students choosing science as the stream, among students enrolled in higher secondary level, is depicted in the bars, along with 95% confidence interval.

Figure 2. Trends in science choice at higher secondary level by caste groups



Note: The figures are estimated using data from multiple rounds of National Sample Survey data on education. Caste wise, the percentage of students choosing science as the stream, among students enrolled in higher secondary level, is depicted in the bars, along with 95% confidence interval.

Figure 3. Gender and caste gaps across the survey rounds



Note: The gaps are estimated by the coefficients from the regression equation (1), including individual and household level controls and district fixed effects. The standard errors are clustered at the household level. The confidence intervals reported are 99, 95 and 90 percent respectively.

Table 1. Summary statistics for the estimation sample

Variable	Year 2007-08		Year 2014		Year 2017-18		Pooled	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Science	0.301	0.459	0.396	0.489	0.379	0.485	0.367	0.482
Commerce	0.126	0.332	0.164	0.370	0.165	0.372	0.156	0.362
Humanities	0.573	0.495	0.440	0.496	0.456	0.498	0.477	0.499
Female	0.411	0.492	0.446	0.497	0.431	0.495	0.432	0.495
OBC	0.423	0.494	0.428	0.494	0.460	0.498	0.439	0.496
SC	0.154	0.361	0.170	0.376	0.184	0.388	0.171	0.377
ST	0.051	0.220	0.078	0.268	0.077	0.266	0.071	0.257
Age	17.004	1.200	16.917	1.176	16.727	1.094	16.861	1.156
Hindu	0.837	0.369	0.834	0.373	0.849	0.358	0.840	0.366
HH Size	5.419	2.384	5.430	2.052	5.229	2.151	5.348	2.174
YoS HoHH <sup>#</sup>	7.165	5.001	6.715	5.047	6.906	4.991	6.895	5.017
MPCE*	2210.3	1628.3	2277.8	1702.1	2361.9	1750.6	2295.4	1705.7
Urban	0.376	0.484	0.323	0.468	0.321	0.467	0.335	0.472
N	6161		9076		14699		29936	

<sup>#</sup>Years of schooling of the head of household.

\*Monthly per capita consumption expenditure (MPCE) in Indian Rupees (INR).

Table 2. Main regression results

VARIABLES	Science			
	(1)	(2)	(3)	(4)
Female	-0.092*** (0.009)	-0.095*** (0.008)	-0.090*** (0.008)	-0.094*** (0.027)
OBC	0.014 (0.012)	-0.023** (0.011)	-0.027** (0.011)	
SC	-0.062*** (0.014)	-0.043*** (0.014)	-0.044*** (0.013)	
ST	-0.150*** (0.017)	-0.060*** (0.018)	-0.069*** (0.018)	
Age		-0.017*** (0.004)	-0.015*** (0.004)	-0.002 (0.010)
Hindu		0.033*** (0.012)	0.030** (0.012)	
Household size		-0.002 (0.002)	-0.001 (0.002)	
YoS HoHH		0.013*** (0.001)	0.014*** (0.001)	
MPCE		0.033*** (0.003)	0.035*** (0.003)	
Urban		0.033*** (0.010)	0.050*** (0.010)	
Year 2014	0.104*** (0.011)	0.118*** (0.010)	0.116*** (0.010)	
Year 2017-18	0.085*** (0.012)	0.089*** (0.010)	0.099*** (0.010)	
Constant	0.350*** (0.011)	0.446*** (0.071)	0.388*** (0.067)	0.430** (0.179)
State FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes
Observations	29,934	29,715	29,714	3,835
R-squared	0.024	0.192	0.251	0.777

Standard errors clustered at the household level are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as an indicator variable for students taking science stream at the higher secondary level. Column 1 shows regression results without any controls. Column 2 shows results with controls and state fixed effects. Column 3 shows the results with controls adding district fixed effects. Column 4 controls for household fixed effects, therefore only those household which have multiple children enrolled in the higher secondary education are considered, and household specific variables such as social group dummies are not identified.



Table 3. Decomposition results

VARIABLES	Decomposition by Gender	Decomposition by Caste		
	(1) Female	(2) OBC	(3) SC	(4) ST
Proportion in Science	0.316*** (0.007)	0.397*** (0.008)	0.323*** (0.011)	0.240*** (0.015)
Proportion in Science for reference group	0.406*** (0.007)	0.379*** (0.008)	0.379*** (0.008)	0.379*** (0.008)
	Gap relative to Male	Gap relative to OC		
Gap	0.090*** (0.009)	-0.018 (0.012)	0.056*** (0.014)	0.139*** (0.017)
Explained	-0.004 (0.003)	0.047*** (0.005)	0.072*** (0.006)	0.067*** (0.007)
Unexplained	0.094*** (0.009)	-0.066*** (0.011)	-0.015 (0.015)	0.072*** (0.018)
<u>Explained:</u>				
Age	-0.006*** (0.001)	-0.006*** (0.001)	-0.001 (0.002)	0.008*** (0.002)
Household size	0.001** (0.001)	0.001* (0.001)	-0.001 (0.001)	0.000 (0.001)
YoS HoHH	-0.001 (0.001)	0.028*** (0.003)	0.048*** (0.005)	0.041*** (0.005)
MPCE	0.003** (0.001)	0.023*** (0.003)	0.027*** (0.004)	0.027*** (0.004)
Urban	-0.001* (0.0003)	0.004*** (0.002)	0.001 (0.003)	-0.001 (0.004)
Year 2014	-0.002** (0.001)	0.002 (0.002)	0.001 (0.001)	-0.002 (0.002)
Year 2017-18	0.001 (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.005*** (0.002)
Observations	29,715	21,851	14,468	13,693

Standard errors clustered at the household level are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as an indicator variable for students taking science stream at the higher secondary level. Column 1 shows decomposition results with between female and male (reference group). In Columns 2 to 4, the reference group is OC. Column 2 shows the decomposition results for the OBC group. Column 3 shows the decomposition results for the SC group. Column 4 shows the decomposition results for the ST group.

Table 4. Results for double jeopardy versus intersectionality analysis

VARIABLES	Science		
	(1)	(2)	(3)
<u>Panel A – Reference group is Male-OC</u>			
Male-OBC	0.013 (0.016)	-0.024 (0.015)	-0.032** (0.014)
Male-SC	-0.073*** (0.020)	-0.053*** (0.018)	-0.056*** (0.018)
Male-ST	-0.172*** (0.022)	-0.088*** (0.023)	-0.098*** (0.022)
Fem-OC	-0.102*** (0.016)	-0.105*** (0.014)	-0.104*** (0.014)
Fem-OBC	-0.087*** (0.016)	-0.126*** (0.015)	-0.125*** (0.015)
Fem-SC	-0.148*** (0.020)	-0.135*** (0.018)	-0.131*** (0.018)
Fem-ST	-0.222*** (0.024)	-0.125*** (0.024)	-0.132*** (0.023)
Constant	0.355*** (0.013)	0.450*** (0.071)	0.394*** (0.067)
<u>Panel B</u>			
Fem-OBC – (Male-OBC + Fem-OC)	0.002	0.003	0.010
SE	(0.022)	(0.020)	(0.019)
P-val	0.918	0.872	0.589
Fem-SC – (Male-SC + Fem-OC)	0.027	0.023	0.028
SE	(0.027)	(0.024)	(0.023)
P-val	0.317	0.339	0.217
Fem-ST – (Male-ST + Fem-OC)	0.052	0.069**	0.070**
SE	(0.032)	(0.030)	(0.028)
P-val	0.103	0.023	0.0134
Other Controls	No	Yes	Yes
State FE	No	Yes	Yes
District FE	No	No	Yes
Round FE	Yes	Yes	Yes
Observations	29,934	29,715	29,714
R-squared	0.024	0.192	0.251

Standard errors clustered at the household level are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as an indicator variable for students taking science stream at the higher secondary level. Column 1 shows regression results without any controls. Column 2 shows results with controls and state fixed effects. Column 3 shows the results with controls adding district fixed effects. Panel A reports the regression results. In Panel A for gender-caste identity groups there are 7 dummy variables included with the reference group being Male-OC. Panel B reports the linear combination of coefficients that test the double jeopardy hypothesis.

Table 5. Regression results with sample selection correction

Variables	Dependent variable: Science		
	(1)	(2)	(3)
Female	-0.092*** (0.008)	-0.088*** (0.008)	-0.119*** (0.036)
OBC	-0.017 (0.011)	-0.025** (0.011)	
SC	-0.030** (0.015)	-0.037*** (0.013)	
ST	-0.046** (0.020)	-0.062*** (0.018)	
Age	-0.014*** (0.004)	-0.014*** (0.004)	-0.002 (0.019)
Hindu	0.017 (0.016)	0.022* (0.013)	
Household size	-0.002 (0.002)	-0.001 (0.002)	
YoS HoHH	0.010*** (0.002)	0.012*** (0.001)	
MPCE	0.030*** (0.004)	0.034*** (0.004)	
Urban	0.032*** (0.010)	0.049*** (0.010)	
Inverse Mills ratio (IMR)	-0.088** (0.043)	-0.041* (0.024)	1.191** (0.518)
IMR × Year 2014	0.024 (0.028)	0.013 (0.023)	-1.137 (0.691)
IMR × Year 2017-18	0.007 (0.028)	-0.001 (0.025)	-3.004*** (0.975)
Constant	0.551*** (0.089)	0.438*** (0.075)	0.621 (1.122)
Observations	29,715	29,714	3,835
R-squared	0.192	0.251	0.782
State fixed effects	Yes	Yes	Yes
Round fixed effects	Yes	Yes	Yes
District fixed effects	No	Yes	Yes
Household fixed effects	No	No	Yes

Bootstrapped standard errors that are clustered at the household level and based on 500 replications are given in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as an indicator variable for students taking science stream at the higher secondary level. The last column controls for household fixed effects, therefore only those household which have multiple children enrolled in the higher secondary education are considered, and household specific variables such as social group dummies are not identified.

## Supplementary Appendix

Table A1. Sample details of the datasets used

Year	July 2007 - June 2008	Jan 2014 – June 2014	July 2017 – June 2018	Pooled
Total Sample				
Households	100,581	65,926	113,756	280,263
Individuals	445,960	310,827	513,366	1,270,153

Note: The table presents the overall sample size for the three rounds of National Sample Survey data on education that we use in the analysis.

Table A2. Net enrollment rates in higher secondary level

Caste	2007-08			2014			2017-18		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
OC	0.308	0.288	0.299	0.446	0.453	0.449	0.502	0.498	0.501
OBC	0.245	0.188	0.219	0.342	0.330	0.337	0.416	0.422	0.418
SC	0.173	0.148	0.162	0.275	0.279	0.277	0.340	0.344	0.342
ST	0.154	0.103	0.131	0.273	0.241	0.258	0.307	0.251	0.281
Total	0.239	0.201	0.222	0.346	0.341	0.344	0.409	0.400	0.405

Note: The enrolment rate is calculated for the age group 16-18 years as the ratio of number of children enrolled in higher secondary to the total number of children in the age group.

Table A3. Regression results for science choice for year 2007-08 (64<sup>th</sup> round)

VARIABLES	Science			
	(1)	(2)	(3)	(4)
Female	-0.071*** (0.015)	-0.096*** (0.013)	-0.099*** (0.014)	-0.020 (0.036)
OBC	-0.003 (0.018)	-0.005 (0.018)	-0.012 (0.018)	
SC	-0.066*** (0.023)	-0.015 (0.022)	-0.017 (0.022)	
ST	-0.165*** (0.029)	-0.080** (0.032)	-0.058 (0.036)	
Age		-0.045*** (0.006)	-0.042*** (0.006)	-0.047*** (0.013)
Hindu		0.017 (0.020)	0.026 (0.020)	
Household size		-0.009*** (0.003)	-0.008*** (0.003)	
YoS HoHH		0.012*** (0.002)	0.013*** (0.002)	
MPCE		0.043*** (0.006)	0.046*** (0.006)	
Urban		0.062*** (0.017)	0.075*** (0.017)	
Constant	0.350*** (0.014)	0.939*** (0.105)	0.865*** (0.104)	1.102*** (0.221)
State FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Observations	6,161	6,155	6,135	981
R-squared	0.013	0.184	0.309	0.791

Standard errors clustered at the household level are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as indicator variable for students taking science course at high school level. Column 1 shows regression results without any controls. Column 2 shows results with controls and state fixed effects. Column 3 shows the results with controls adding district fixed effects. Column 4 controls for household fixed effects, therefore only those household which have multiple children enrolled in the higher secondary education are considered, and household specific variables such as social group dummies are not identified. Note that with household fixed effects there are not enough samples (households with variation in gender and children enrolled in higher secondary), which could affect the credibility of the estimates reported.

Table A4. Regression results for science choice for year 2014 (71<sup>st</sup> round)

VARIABLES	Science			
	(1)	(2)	(3)	(4)
Female	-0.107*** (0.016)	-0.112*** (0.015)	-0.114*** (0.014)	-0.112** (0.047)
OBC	0.026 (0.020)	-0.013 (0.020)	-0.031* (0.019)	
SC	-0.066*** (0.025)	-0.044* (0.024)	-0.047** (0.023)	
ST	-0.137*** (0.028)	-0.054* (0.030)	-0.067** (0.029)	
Age		-0.023*** (0.007)	-0.021*** (0.006)	-0.001 (0.019)
Hindu		0.013 (0.021)	0.004 (0.020)	
Household size		0.003 (0.004)	0.003 (0.004)	
YoS HoHH		0.014*** (0.002)	0.014*** (0.002)	
MPCE		0.030*** (0.006)	0.032*** (0.007)	
Urban		0.025 (0.016)	0.055*** (0.016)	
Constant	0.455*** (0.017)	0.647*** (0.120)	0.633*** (0.116)	0.448 (0.325)
State FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Observations	9,076	9,012	9,008	1,292
R-squared	0.022	0.180	0.316	0.780

Standard errors clustered at the household level are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as indicator variable for students taking science course at high school level. Column 1 shows regression results without any controls. Column 2 shows results with controls and state fixed effects. Column 3 shows the results with controls adding district fixed effects. Column 4 controls for household fixed effects, therefore only those household which have multiple children enrolled in the higher secondary education are considered, and household specific variables such as social group dummies are not identified.

Table A5. Regression results for science choice for year 2017-18 (75<sup>th</sup> round)

VARIABLES	Science			
	(1)	(2)	(3)	(4)
Female	-0.091*** (0.017)	-0.083*** (0.014)	-0.078*** (0.013)	-0.124*** (0.044)
OBC	0.013 (0.021)	-0.042** (0.019)	-0.039** (0.017)	
SC	-0.053** (0.024)	-0.062*** (0.023)	-0.055** (0.022)	
ST	-0.153*** (0.027)	-0.060** (0.029)	-0.073*** (0.026)	
Age		0.004 (0.007)	0.002 (0.006)	0.042** (0.018)
Hindu		0.063*** (0.021)	0.046** (0.019)	
Household size		-0.000 (0.004)	0.001 (0.003)	
YoS HoHH		0.013*** (0.002)	0.013*** (0.002)	
MPCE		0.030*** (0.004)	0.030*** (0.004)	
Urban		0.023 (0.017)	0.031** (0.015)	
Constant	0.433*** (0.017)	0.168 (0.118)	0.198* (0.107)	-0.250 (0.310)
State FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Observations	14,697	14,548	14,546	1,562
R-squared	0.018	0.230	0.354	0.775

Standard errors clustered at the household level are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as indicator variable for students taking science course at high school level. Column 1 shows regression results without any controls. Column 2 shows results with controls and state fixed effects. Column 3 shows the results with controls adding district fixed effects. Column 4 controls for household fixed effects, therefore only those household which have multiple children enrolled in the higher secondary education are considered, and household specific variables such as social group dummies are not identified.



Table A6. Regression results by rural and urban regions

VARIABLES	Rural				Urban			
	(1)	(2)	(3)	Science Choice (4)	(5)	(6)	(7)	(8)
Female	-0.105*** (0.012)	-0.109*** (0.011)	-0.106*** (0.011)	-0.112*** (0.036)	-0.076*** (0.015)	-0.067*** (0.012)	-0.061*** (0.012)	-0.050 (0.037)
OBC	0.006 (0.015)	-0.047*** (0.015)	-0.048*** (0.015)		0.069*** (0.017)	0.024 (0.015)	0.012 (0.016)	
SC	-0.040** (0.018)	-0.067*** (0.018)	-0.065*** (0.017)		-0.044* (0.023)	0.007 (0.022)	-0.010 (0.021)	
ST	-0.133*** (0.020)	-0.079*** (0.020)	-0.085*** (0.022)		-0.109*** (0.037)	-0.052 (0.041)	-0.044 (0.034)	
Age		-0.012** (0.005)	-0.007 (0.005)	-0.009 (0.014)		-0.030*** (0.006)	-0.030*** (0.006)	0.014 (0.015)
Hindu		0.019 (0.017)	0.013 (0.017)			0.045*** (0.017)	0.044*** (0.016)	
HH size		0.002 (0.003)	0.002 (0.003)			-0.011*** (0.003)	-0.008** (0.003)	
YoS HoHH		0.011*** (0.001)	0.011*** (0.001)			0.018*** (0.001)	0.017*** (0.001)	
MPCE		0.041*** (0.009)	0.037*** (0.009)			0.028*** (0.003)	0.034*** (0.003)	
Year 2014	0.136*** (0.014)	0.147*** (0.013)	0.144*** (0.013)		0.059*** (0.018)	0.072*** (0.016)	0.073*** (0.016)	
Year 2017-18	0.103*** (0.014)	0.119*** (0.013)	0.135*** (0.013)		0.069*** (0.020)	0.040** (0.016)	0.045*** (0.016)	
Constant	0.293*** (0.014)	0.343*** (0.093)	0.271*** (0.089)	0.546** (0.235)	0.419*** (0.017)	0.698*** (0.101)	0.682*** (0.100)	0.160 (0.257)
State FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes	No	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	16,025	15,910	15,907	2,020	13,909	13,805	13,802	1,815
R-sq	0.029	0.176	0.260	0.772	0.019	0.208	0.289	0.789

Standard errors clustered at the household level are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as indicator variable for students taking science course at high school level. Column 1 to column 4 pertain to Rural subsample and columns 5 to 8 pertain to Urban subsample. Column 1 & 5 show regression results without any controls. Column 2 & 6 show results with controls and state fixed effects. Column 3 & 7 show the results with controls adding district fixed effects. Column 4 & 8 control for household fixed effects, therefore only those household which have multiple children enrolled in the higher secondary education are considered, and household specific variables such as social group dummies are not identified.

Table A7. Non-linear regression of science choice (marginal effects)

VARIABLES	Science			
	(1)	(2)	(3)	(4)
	<u>Probit</u>		<u>Logit</u>	
Female	-0.092*** (0.009)	-0.095*** (0.008)	-0.093*** (0.009)	-0.095*** (0.008)
OBC	0.013 (0.011)	-0.019* (0.011)	0.013 (0.011)	-0.019* (0.011)
SC	-0.062*** (0.014)	-0.041*** (0.014)	-0.062*** (0.015)	-0.040*** (0.014)
ST	-0.159*** (0.019)	-0.061*** (0.019)	-0.162*** (0.020)	-0.063*** (0.020)
Age		-0.017*** (0.004)		-0.017*** (0.004)
Hindu		0.037*** (0.012)		0.037*** (0.013)
Household size		-0.002 (0.002)		-0.001 (0.002)
YoS HoHH		0.013*** (0.001)		0.013*** (0.001)
MPCE		0.035*** (0.004)		0.037*** (0.004)
Urban		0.030*** (0.010)		0.028*** (0.010)
Year 2014	0.105*** (0.012)	0.116*** (0.010)	0.105*** (0.012)	0.116*** (0.011)
Year 2017-18	0.087*** (0.012)	0.088*** (0.010)	0.087*** (0.012)	0.089*** (0.011)
State FE	No	Yes	No	Yes
District FE	No	No	No	No
Round FE	Yes	Yes	Yes	Yes
Observations	29,934	29,715	29,934	29,715

Standard errors clustered at the household level are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as an indicator variable for students taking science courses at the high school level. Columns 1 & 2 show results of Probit model, while, columns 3 & 4 show results of Logit model. Columns 1 & 3 show regression results without any controls. Column 2 & 4 show results with controls and state fixed effects.

Table A8. Intersectionality results – alternative specification using interaction terms

VARIABLES	Science		
	(1)	(2)	(3)
Female	-0.102*** (0.0157)	-0.105*** (0.0142)	-0.104*** (0.0136)
OBC	0.013 (0.016)	-0.024 (0.015)	-0.032** (0.014)
SC	-0.073*** (0.020)	-0.053*** (0.018)	-0.056*** (0.018)
ST	-0.172*** (0.022)	-0.088*** (0.023)	-0.098*** (0.022)
Fem × OBC	0.002 (0.022)	0.003 (0.020)	0.010 (0.019)
Fem × SC	0.027 (0.027)	0.023 (0.024)	0.028 (0.023)
Fem × ST	0.052 (0.032)	0.069** (0.030)	0.070** (0.028)
Age		-0.017*** (0.004)	-0.015*** (0.004)
Hindu		0.033*** (0.012)	0.030** (0.012)
Household size		-0.002 (0.002)	-0.001 (0.002)
YoS HoHH		0.013*** (0.001)	0.014*** (0.001)
MPCE		0.033*** (0.003)	0.035*** (0.003)
Urban		0.033*** (0.010)	0.050*** (0.010)
Year 2014	0.103*** (0.011)	0.118*** (0.010)	0.116*** (0.010)
Year 2017-18	0.085*** (0.012)	0.089*** (0.010)	0.099*** (0.010)
Constant	0.355*** (0.013)	0.450*** (0.071)	0.394*** (0.067)
State FE	No	Yes	
District FE	No	No	Yes
Round FE	Yes	Yes	Yes
Observations	29,934	29,715	29,714
R-squared	0.024	0.192	0.251

Standard errors clustered at the household level are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variable 'Science' is defined as an indicator variable for students taking science stream at the higher secondary level. Column 1 shows regression results without any controls. Column 2 shows results with controls and state fixed effects. Column 3 shows the results with controls adding district fixed effects.