Labor Market Effects of High School Science Majors in a High STEM Economy*

Tarun Jain† Abhiroop Mukhopadhyay‡ Nishith Prakash§
Raghav Rakesh¶

April 18, 2018

This paper estimates the labor market effects of high school major choices in India, a country with a very high proportion of STEM majors among graduates. In many countries including India, decisions to undertake STEM fields of study are dictated by high school major choices; hence estimating the impact of such pre-college choices on earnings is important to understand why India produces a large number of STEM graduates. Using a representative household survey that collected unique information on high school major choices, we estimate that those who have studied science in High School earn 22 percent more than those who studied business and humanities, even after we control for observed markers of ability: *academic performance* and *English language fluency*. Our point estimates are stable when we add additional covariates and satisfy Altonji et al. (2005)-Oster (2017) bound analysis. Studying science is also correlated with subsequent education progression and knowledge of computers. Moreover, our results indicate that some disadvantaged groups may be left out of reaping the science premium. Finally, using a primary data on high school students in two large states (Bihar and Andhra Pradesh) of India, we explore cognitive and non cognitive factors that may correlate with science choice and find that science students have positive personality traits, have higher cognitive reflection scores and are more ambiguity averse.

JEL Codes: I23, I26, J24
Keywords: High-School Major Choices, Economic Returns, Science, STEM, India

*We thank Mehtabul Azam, Eric Edmonds, Caroline Hoxby, Karthik Muralidharan, Stephen Ross, Duncan Thomas, Sarah Turner and seminar and conference participants at the 13^{th} Annual Conference on Economic Growth and Development for their helpful comments and suggestions. We thank Revathy Suryanarayana for outstanding research assistance. Primary data collection sponsored by the International Growth Centre Bihar and the Planning and Policy Research Unit at ISI Delhi, and approved by IRB at Indian School of Business, University of Connecticut and the ethics committee at Indian Statistical Institute.

†Indian School of Business. Email: tarun_jain@isb.edu.
‡Indian Statistical Institute and IZA. Email: abhiroop@isid.ac.in.
§University of Connecticut, IZA, HiCN, and CReAM. Email: nishith.prakash@uconn.edu.
¶Indian School of Business and Indian Statistical Institute. Email: raghavrakesh92@gmail.com.

"[Science] is more than a school subject, or the periodic table, or the properties of waves. It is an approach to the world, a critical way to understand and explore and engage with the world, and then have the capacity to change that world..."

- President Barack Obama, March 23, 2015.

1 Introduction

India is second, only after China, in educating college graduates specializing in Science, Technology, Engineering and Mathematics (STEM).[1] Among graduates in India, 35% are STEM majors while 53% are humanities in 2012.[2] By 2030, China and India are likely to account for more than 60% of the OECD and G20 STEM graduates. Despite the success of these countries in producing STEM graduates, and the attempt of other countries to follow, the labor market consequences of STEM education are unclear. Understanding the influence of STEM on eventual earnings as well as the pathways that enable the earnings is an important question for both researchers and policymakers. Filling this gap will allow a better understanding of the process through which science-heavy education translates into livelihoods and earnings, as well as help design policies to encourage students and administrators to pursue the most productive educational paths.

This paper estimates returns to studying science, as proxy for STEM, using a nationally representative data from India. This is a useful setting to investigate this question because the Indian education structure allows students to specialize in either science, business or humanities at the higher secondary stage of school. A large share of students opt for science at this stage – among working age men in urban India, approximately 25.8% had studied science, 24.6% had studied business and 49.6% had a background in humanities. This intense, focused education in science subsequently guides future decisions in college and the workforce.

[1] China produces 4.7 million STEM graduates, closely followed by India at 2.6 million, and the United States at 568,000 (World Economic Forum, 2016).

[2] Organisation for Economic Co-operation and Development (OECD) database, National Statistics websites for China and India.

In contrast, education systems in North America and the United Kingdom do not allow such specific focus, which makes isolating the influence of science difficult. In addition, India is a large producer of science, engineering and technology graduates, where 35% in the graduate pool are STEM graduates OECD (2015). Our paper estimates both the returns to science in this setting, while illustrating the pathways such as complementarities with English and computer skills, the role of parental and social background, increased post-secondary education, and types of job, that translate science education into higher earnings.

Estimating the returns to science faces a number of challenges. From an empirical perspective, estimating the causal impact of high-school major choices on earnings is not straightforward due to endogeneity of the major choice variable in the earnings estimation. Omitted variables such as ability, language and communication skills or labor market conditions could bias estimates of the relationship between the choice of major and earnings. In order to address these biases, we exploit the richness of the India Human Development Survey (IHDS) data that allows us to control for ability using performance in the tenth grade exam. Similarly, the IHDS also allows us to control for English-language fluency which could be correlated with the ability to do well in science, admission tests, and job interviews. We add state and district characteristics that control for labor market conditions, and age, marital status, caste and religion dummies to control for individual demographic characteristics.[3] These controls allow us to report tight correlations between major choice and labor market earnings.

We find several interesting results. First, studying science yields sizable returns in the labor market. After controlling for proxies for ability, English-language ability, geography and various demographics, we find that in urban India, mean annual earnings are 22% higher

[3]Despite controlling for these omitted variables, subject choice could signal status or demonstrate grit, factors that could simultaneously influence earnings but are not adequately captured by our measure of ability. However, only experimental variation where students are randomly assigned to high school majors will overcome the remaining identification challenges, something that is extremely challenging in this context. Wiswall and Basit (2015) study the determinants of college major choice using an experimentally generated panel of beliefs, obtained by providing students with information on the true population distribution of various major-specific characteristics. They find that while expected earnings and perceived ability are a significant determinant of major choice, heterogeneous tastes are the dominant factor in the choice of major.

for men who study science in high school relative to men who study business and humanities. Even after controlling for parental education, we find the returns to be 21% higher for men who study science. Results from quantile regressions show that the returns are more or less similar at all points of the wage distribution except the highest 1 percent: returns to studying science are as high as 37% for the 99th percentile income earners.

Second, heterogeneity analysis by ability suggest differential returns to studying science relative to business and humanities at *all levels of ability*. We find higher relative marginal returns to studying science *only* when individuals have at least moderate level of English-language fluency, which suggests a strong complementarity between English-language fluency and science. Moreover, we find a strong complementarity between knowing computers and studying science, suggesting that computer fluency is important for translating science knowledge into higher earnings. Further heterogeneity analysis suggests that the returns to science are concentrated among those who do not have professional degrees.[4] We also find that the returns do not accrue, on average, to disadvantaged Scheduled Caste (SC) and Scheduled Tribe (ST) communities.

Third, in order to understand channels, we find that those with science in high school are likely to have more years of education, more likely to have at least an undergraduate education, and more likely to complete professional courses. We find no average effect of studying science on employment type (private employee, public employee or businessman), though people with low scholastic ability are more likely to get public employment with science. When looking the impact of studying science on the income, conditional on one of the employment types, we find that the highest returns from studying science accrue to those who are businessmen and have a high scholastic scores.

Given the lack of an experimental research design with exogenous variation in the choice to study science, we investigate the extent to which omitted variables can affect our results. Using methods developed by Altonji et al. (2005) and Oster (2017), we find that our estimate

[4] Professional courses include engineering, medicine, management, accountancy and law

interval bounds do not include the null, and that the estimated returns to studying science are robust to potentially large selection on the basis of unobservables.

While the large data set allows us to give representative results for urban males in the country, it lacks in measures of cognitive and non-cognitive skills. While the robustness procedure described above may alleviate concerns about these omitted factors overturning the main result, we know that such skills play a role in the labor market (Heckman and Kautz (2012)). Using a primary survey of 558 students in grade 12 (the last year of high-school) across 44 schools in two large states of India (Andhra Pradesh and Bihar), we test the extent to which cognitive and non-cognitive skills differ across students who choose science and non-science majors. We find that science students have positive personality traits and higher cognition reflection test scores that reflect more patient behavior, but are more ambiguity averse. These results are consistent with science students postponing irreversible choices till they know more about potential downstream educational and labor market options, combined with the patience to wait.

This paper contributes to a number of strands of the literature. First, the literature on returns to STEM majors in the United States has focused on understanding why a large share of college students drop out of STEM major, with only a handful of papers estimating the returns to studying STEM major in college. Among these, Black et al. (2015) examine the relationship between courses that provide STEM training in high school and later labor market success as measured by wages as well as employment in a STEM occupation. They find that mathematics courses are an important predictor of labor market success, even after controlling for cognitive test scores and fixed high school characteristics. Similarly, Altonji et al. (2012) show that the wage gap between electrical engineers and general education majors is within two percentage points of the gaps between college and high school graduates. Beyond STEM, a larger literature documents differences in earnings across majors for college graduates using quasi-experimental variation in student assignment to different majors.[5] Arcidiacono (2004) finds that mathematics ability is important for labor market returns and for sorting into particular majors, and that after controlling for this selection, students who select natural science and business majors receive large financial returns. Our paper is the first within this topic drawing on data from a developing country, where the nature of education as well as the structure of labor market might qualitatively change the returns to STEM education.

We also add to the literature on returns to human capital by estimating returns to "stream choices" in a developing country context. The literature on the returns to education largely focuses on the returns to years of schooling (Card (1999)), while ignoring the role of stream choices, language skills, computer skills and other dimensions of schooling. Notable exceptions that unpack the impact of education content include Munshi and Rosenzweig (2006), Chakraborty and Kapur (2009), Azam et al. (2013) and Jain et al. (2016). In particular, Munshi and Rosenzweig (2006) estimate the returns to studying in English-medium school and find significant positive returns for both men and women. Chakraborty and Kapur (2009) exploit a policy change in the Indian state of West Bengal which changed the medium of instruction in primary school from English to Bengali (native language) and finds a significant negative impact on wages for the exposed cohorts. Similarly, Azam et al. (2013) use IHDS (2005) to find that after controlling for a large number of individual and environmental factors, wages are on average 34% higher for men who speak fluent English relative to men who do not.

Finally, our paper relates to the literature on determinants of the stream and occupational choices, a classic research question in social sciences. This literature focuses on two sets of relationships: occupational choices and future expected earnings, and college major choices and occupational choices. Several papers including Grogger and Eide (1995); Brown and Corcoron (1997); Weinberger (1998); and Gemici and Wiswall (2014) have documented

[5] See Altonji et al. (2012), Altonji et al. (2015), Daymont and Andrisani (1984), Grogger and Eide (1995), Hastings et al. (2013), Kirkeboen et al. (2015), James et al. (1989), Loury (1997), Loury and Garman (1995) and Gemici and Wiswall (2014).

that post-secondary field choice is an important determinant future expected earnings, and most importantly, college major choices can provide insights to understand long-term changes in inequality and earnings differences by gender and race.

2 Backgound

In the Indian education system, students receive ten years of basic education supplemented by two years of senior secondary education and three years to five years of higher education (MHRD, Government of India (1998)). The objective of the first ten years is to provide a well-rounded, non-selective general education to all students. The two years of senior secondary education allows students to specialize, while preparing them for higher education. In contrast to the United States, students in India must choose at this stage from three standard majors or specialization in the academic streams: Science, Business and Humanities.[6] The duration of higher education varies from three to five years depending on the course. Bachelor in Arts and Science programs last three years, technical courses are four years, and medicine and architecture take five years. Thereafter, students opt for further higher education or enter the labor force.

A unique feature of the Indian education system is the deterministic role of high school major choices on college majors, where these pre-college major choices are largely irreversible. Students who choose science as a high school major are the only ones eligible to study STEM courses in college. However, they are also eligible to pursue various non-STEM courses. Conversely, students who chose business or humanities in high school are eligible to *only* pursue non-STEM courses in college. Therefore, high school major choices directly effect the set of courses one can pursue after high school, and is considered to be a critical first step in long-term career paths.

[6] The State or All-India Boards of Secondary Education determine curriculum at the higher secondary level. The curriculum that a particular school follows will be determined by the State or All-India board to which it is affiliated. Those schools affiliated with the new-Delhi-based Central Board of Secondary Education, for example, will follow its curriculum and offer All India Senior School Certificate Examinations administered by the Board.

In theory, any student who obtains a passing grade, i.e., 30-35% out of a maximum of 100% (depending on the examination board), in the Secondary School Certificate (SSC) examination (grade 10) can be admitted to the 11th grade.[7] However, in practice, the eligibility criteria in terms of grade 10 performance is higher for science stream as compared to Business or Humanities [**reference?**]. This could be partly attributed to a demand-supply mismatch in number of seats in high schools, and partly to schools' beliefs that students who study science have higher ability as measured by their grade 10 performance [**reference?**]. In our estimation sample of individuals age 25-65 residing in urban India, approximately 25.8% studied Science, 24.6% studied Business, and the remaining 49.6% studied Humanities. At the country level, India produces approximately 2.6 million STEM graduates each year and 35% of the Indian graduates are STEM majors (World Economic Forum, 2016).

There are multiple potential factors why students take science in high school in India, apart from heterogeneity in tastes. One major factor is possibly that the science stream and associated career paths are more prestigious. Another possible determinant could be that technical and professional courses following science stream can get student a secured job immediately after college without pursuing any other degree, which is not similar to non-science streams. So, immediacy of the job is another factor which is driving students to science stream in high school. Moreover, the pool of jobs available after STEM courses is much wider than after a non-STEM program as science students can pursue various non-STEM courses whereas it is not possible otherwise. So, there is higher probability for a science student to get a job.

In India, there are large monetary costs involved in pursuing science major. One major cost incurred by students studying science is large monetary investments in private tuitions to prepare for entrance examinations to elite engineering and medical colleges. Enrollment in these coaching institutes might start as early as grade 9 (sometimes even grade 7) for a four-year integrated program. Also, anecdotal evidence suggests that students in science

[7] In India, students must pass the SSC examination to be eligible for further schooling and a better score in this exam enables students to attend better schools.

stream take the highest number of private coaching for individual subjects to score well in the higher secondary examination (grade 12), even if they are not preparing for these competitive examinations.[8] Another flourishing trend among science students is that they leave the native place and migrate to urban coaching hubs within the state or famous cities in another state which have established themselves as national coaching hubs over the years. This further involves a non-pecuniary cost of staying away from family at a young age along with pecuniary costs of coaching fees and basic subsistence payments like payments for food and lodging during the entire duration of coaching.

It is widely believed that there are higher expected returns to studying science in high school, but we are not aware of any estimate. The IHDS data suggests that the earnings of individuals who chose Science as a high school major in 10+2 are significantly higher relative to those from Business or Humanities streams. Similarly, students who choose Business as a high school major earn more than those who studied Humanities. In this paper, we provide first evidence on the distribution of high school major choices among the adult population in urban India and subsequently estimate the labor market returns.

3 Data

The main empirical analysis uses the India Human Development Survey (IHDS) data collected in 2011-12 by the National Council of Applied Economic Research. IHDS is a nationally representative, multi-topic household survey covering 42,152 households across India.[9] Uniquely among nationally representative household surveys, the IHDS collects data on individuals major choices in high school and their current earnings. In addition, the survey

[8] We are not aware of any paper that estimates such costs for students studying science, however Azam (2016) reports that the average cost of private tutoring to those who took private tutoring in 2007-08 is about 42.7% of total private education expenditure, which is about 16.5% of household per capita expenditure. This jumps to approximately 40% of total private expenditure on education at the secondary and senior secondary level.

[9] The survey covered all the states and union territories of India except Andaman and Nicobar Islands and Lakshadweep islands, which together account for less than 0.05% of India's population. For data analysis, we use IHDS design weights to obtain nationally representative statistics.

also reports variables associated with demographic characteristics, ability, English-language fluency, computer skills, educational achievement and occupational outcomes.

Our sample consists of urban males aged 25 to 65 who have completed at least secondary schooling (grade 10), made a stream choice, and report information on both major choices and earnings. We do not include women in the analysis because of low female labor force participation (Klasen, S. and Pieters, J. (2015)).[10] We do not consider rural residents because measuring agricultural and in-kind income is difficult. These restrictions yield 4,763 men in the sample. Most of them report working for wage/salaried employment, business, and self-employed. Only 4 percent are reported to be unemployed. Since they earn zero income, we drop them as we use log of earnings as our dependent variable for the main results.[11]

In the IHDS, household level enterprise profit is reported along with the labour time contribution in the enterprise of household members. We use this information to calculate earnings for business/self-employment.[12] Using the household's net enterprise profit (already net of costs), we apportion the amount of the net profit based on the individual's share of the total time spent by household members on enterprise activities. Such apportioning avoids the need for selection models, for which identifying variables are difficult to find.

Further, we do not look at wage (earnings) rate but instead focus on annual earnings. This incorporates both the wage rate as well as the number of hours worked in the year. Given that labour is typically inelastically supplied by most male adult members of households [**needs citation support.**] in developing countries like India, the actual amount of work is likely to reflect demand for labour. The demand for labour is an important part of the earning payoff for an individual. For example, while public salaried employment may not offer the highest wage rate in the labour market, the fact that most public employees are assured work throughout the year, ensures larger earnings returns from such jobs.

[10] https://www.nytimes.com/2015/08/24/opinion/why-arent-indias-women-working.html

[11] An alternate specification where the dependent variable is in levels gives us similar results even if we include the unemployed.

[12] Typical regressions that calculate Mincerian returns in the context of India include *only* wage employees. This is dictated by the lack of earnings data in the employment datasets of the National Sample Survey, the most commonly used dataset for estimating earnings in India.

We report descriptive statistics for our dependent variables in Table 1. The first row reports the mean annual earnings are Rs. 178,330 ($2,830). Respondents who have completed 10^{th} grade have, on average, 3.86 years of further education, representing completion of high school and some college. In our sample, 26% report to be working in public employment, 25% in private employment and 27% in business employment.

Table 1 reports descriptive statistics for various control variables used in the estimation. Twenty five percent of the sample studied science in high school. Approximately 32% received a first division ($> 60\%$ score), 57% received a second division ($50\% <$ score $< 60\%$), 12% received a third division ($40\% <$ score $< 50\%$), and 12% repeated a grade. Similarly, 35% of the sample speaks fluent English, 48% speak English less fluently while the remaining 17% cannot speak any English. Among the demographic variables, the average age is slightly less than 40 years, 83% men are married, 33% belong to Other Backward Classes, 12% are Scheduled Castes, 3% are Scheduled Tribes, 8% are Muslims and 3% are Christians.[13]

Given the background of those surveyed, is there a difference in earnings between those studying science and those studying other majors? Figure 1 plots the distribution of log earnings by major choice showing that the distributions are different, with the mean log earnings for science students higher than students from other majors. The mean earnings for science students is Rs. 224,194 ($3,558) while that of students from other majors is Rs. 156,000 ($2,476). This difference remains even while conditioning on the scholastic ability of the individual although the two density functions are far closer for those with first division as compared to those with lower divisions.

Finally, in order to understand other correlates of studying science major, we conducted a primary survey in six districts across Bihar and Andhra Pradesh states in India. The survey was conducted at the beginning of the academic calendar in the months of May, June and July in 2017. The six districts covered by the survey are Patna, Bhagalpur and Sitamarhi

[13]The Scheduled Castes and Scheduled Tribes are historically disadvantaged minorities recognized by the Constitution in India. The Government of India classifies approximately 41% of the country's population as Other Backward Class (OBC) who are socially and educationally disadvantaged.

in Bihar, and Vijayawada, Kurnool and Srikakulam in Andhra Pradesh. The survey was conducted only in districts towns of the mentioned districts. The schools in these towns were randomly chosen by stratifying on private versus public management. Students of grade 12 across the three streams (Science, Business, and Humanities) were then chosen at random from the school lists and interviewed at their home. Separately, we also interviewed one of the parents. We only surveyed students who were currently co-residing with either father or mother, or both at their local residence.[14]

Apart from the standard variables, the main purpose of the survey was to collect various measures of cognitive and non-cognitive skills that is typically not collected in most surveys in developing countries.[15] The survey consists of 461 students and their parents across 38 public and private schools in Bihar and 424 students and their parents across 19 public and private schools in Andhra Pradesh. The summary statistics for the survey sample is presented in the Tables 12 and 13.

4 Empirical Analysis

4.1 Specification

We use the 2011-12 round of the IHDS data with one observation per individual, and estimate the returns to studying science major in high school using the following regression specification.

$$y_i = \beta_0 + \beta_1 Science_i + \beta_2 \mathbf{X}_i + \lambda_d + \epsilon_i \tag{1}$$

In equation (1), the main outcome of interest y_i is log earnings of an individual i residing in district d. The variable $Science_i$ is 1 if the individual studied science in high school,

[14]It is not uncommon in India for students to reside with their relatives (or local guardians), especially if parents live in rural areas lacking good schools. We did not survey such students since our goal was to survey both the student and the parent, and it is less likely for relatives (or local guardians) to influence major choice or career decisions of students.

[15]We describe these variables in Appendix 8.

and 0 otherwise. The coefficient of primary interest is β_1, which is percent increase in earnings associated with studying science in high school. We add a vector of control variables, represented by $\mathbf{X}_i$, which includes a measure of ability, represented by indicator variables for whether the individual obtained first, second and third division in the grade 10 exam.[16] Students with higher ability are more likely to study science in high-school as well as have better jobs, leading to an upward bias on the estimate of the return to studying Science if a measure of ability is omitted.[17] IHDS data allows us to control for ability as it reports individual's performance on the secondary school certificate (SSC) examination.[18] In order to control for ability among the less educated, we add whether the individual ever failed or repeated a grade. In addition, we also control for average household education (excluding the respondent) to proxy for household level ability.[19]

Fluency in English directly effects labor market returns (Azam et al., 2013), so $\mathbf{X}_i$ includes measures for self-reported English fluency, represented by indicator variables for "very fluent", "little fluent" and "not fluent". Third, we add rich set of control variables for individual age, marital status, religious and social group. The specification includes district fixed-effects (λ_d) which controls for all geographic, economic and social factors that are common to all individuals within a district. Finally, the term ϵ_i represents *i.i.d.* unobserved factors that might influence earnings.

In addition to log earnings, we estimate equation (1) for a number of follow-on outcome variables. These are variables representing human capital achievement (specifically, years of

[16]This is similar to controlling for aptitude test scores to address the ability bias when estimating the returns to schooling.

[17]According to the authors calculation using the estimation sample, 39.31% of students who receive division I (higher ability), 20.48% of students who receive division II, and only 10.22% of students who receive division III choose to study Science in grade 10.

[18]SSC is a standardized exam developed by the board of education and is taken at the end of grade 10, and the passing categories, from highest to lowest level of distinction are I, II and III division. Majority of students study in schools where exams are conducted by the state board making the divisions comparable within states.

[19]These control for advantages of belonging to educated households. As a robustness, we also control for parental education, a traditional control to proxy for ability in the returns to education literature (Card, 1999).

schooling, whether the individual completed graduate education, and whether the individual completed professional education) and employment (in particular, public sector salaried employment, private sector salaried employment, and employment in business, as well as income associated with these categories of employment).

4.2 Main results on earnings

Table 2 reports findings from estimating equation 1, sequentially introducing controls in Columns 1 through 5. The main result in Column 5, after including all control variables, is that studying Science in Column 5 is associated with 21% higher earnings ($p < 0.01$). The magnitude of this coefficient is comparable to the influence of *"Fluent English"* skills, (+36.0%, consistent with estimates reported by Azam et al. (2013)), indicating the importance of high school curriculum on adult earnings. Also important is household education, with a year increase in average education of the other household members being associated with three percent greater earnings [20].

Examining heterogeneity in the results helps to determine the pathways through which transmission from science education to earnings occur. We first examine heterogeneity by earnings quintile, which reveals the relative importance of science education for students at the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} and 99^{th} percentiles of the earnings distribution. Studying science has a comparable uniform significant influence on earnings at all these points of the wage distribution, except for those at the top 1 percent. At the 99^{th} percentile, we find that returns to studying science in high school is 37 percent. It is well known that the returns to education are convex for India and many developing countries. Our results throw further light on what drives this convexity by highlight that stream choice is correlated with the highest incomes.

[20] Table 11 reports finding on estimating equation 1 after including the unemployed population, which is not included in our final sample. In column 5, after including all control variables, studying science is associated with 15% higher earnings ($p < 0.05$). This suggests robust returns to studying science after including unemployed population. The earnings of unemployed people was coded as 1 so that log(earnings) becomes 0.

Returns to science education might be conditioned by students' ability, with higher ability students potentially more able to translate knowledge of science into greater earnings. Table 4 examines this empirically by dividing the sample among those who received a first division ($> 60\%$ grade) versus a second or third division ($40\% <$ grade $< 60\%$) scores in tenth grade, and reporting separate results from estimating equation (1). We find that the point estimate of returns to science are higher for students with first (+0.25% greater earnings) versus lower (+0.19% greater earnings) division scores in tenth grade. Although the two estimates are not statistically different from each other, these findings, along with those from the quintile regressions, are consistent with science education complementary with ability, with the greatest marginal value for the most capable students and employees.

We also explore the complementarity of science with other skills, specifically spoken English and computer fluency. Such complementarities might be particularly important in the labor market, where the structure of jobs might dictate the returns to skills. If science jobs also require extensive communication with others, especially in the business world where language skills are important, then the returns to science might be influenced by English fluency. Conversely, if STEM careers require expertise in science with communications handled by other employees, then students with science proficiency could obtain high returns independent of their language skills. For similar reasons, the value of STEM education could depend on knowledge and fluency with computers.

Without precisely defining the production function, the empirical exercise offers insight into the complementarities between science education and English language and computer skills. Panel A of Table 5 finds that the only returns to science accrue when an individual knows English. The earnings returns are 28% greater with fluency in spoken English ($p < 0.01$), and 19% higher with little English ($p < 0.01$). The returns to science are statistically indistinguishable from zero without English, regardless of ability measured by tenth grade scores, indicating the critical role of English language skills in complementing STEM education in the job market. Mirroring these results are the findings associated with computer skills in Panel B. Science education is associated with high returns (+31% for first division students, $p < 0.01$; +19% for second and third division students, $p < 0.05$), but only when the respondent was proficient in computers. Returns are significantly lower (7%, $p < 0.10$) for students who report that they are not proficient in computers. Collectively, these findings point to the critical role of communication and technical skills in operationalizing the returns to science education.

The returns to education literature for India has shown that returns are highest to market oriented courses like technical education (Duraiswamy (2002)). Such courses typically command higher wages as compared to "general" university education. However, technical courses are only one subset of many courses that are market oriented: courses like MBAs, Law, Chartered accountancy (CAs) are also lucrative on the labour market. In fact, typically many students follow up on their engineering degrees with MBAs in the early years of their career to increase their wages. We refer to such courses aimed at the market: technical courses (Engineering, Medicine), Law, MBA and CAs as professional degrees. Non-technical professional courses account for 54% of professional degrees. A feature of students who have completed such professional degrees is that their skills sets are in sync with the market. It is therefore interesting to ask whether science education has complementarity with market oriented skills.[21] Interestingly, we find that science education has no impact when comparisons are made among those who have completed professional degrees. This is not because completing professional degrees correlates perfectly with studying science: among those with professional degrees, science and non-science school majors are almost equally represented, since 43.6% of those with professional degrees did not study science in high school. The relative insignificance of science is true across all ability levels. In contrast, we find studying science has much larger return among those without such market skills: that is those who did not complete a professional degree. This result is equally robust across all ability levels. This points, out that if one can get a market oriented degree, the advantages of studying

[21] It is equally important to ask whether science students are more likely to get market oriented skills. We investigate that in the next section.

science are nullified, but for those in the market without market oriented education, studying science at the school level plays a very important role in the labour market. This is an important point to consider as only 5.5% of our sample has a professional degree.

We next examine how the social environment, represented by social group and parent education, influences the value of studying science. Socially privileged individuals might benefits disproportionately more from STEM education, since they might have access to job and commercial opportunities required to convert their education into higher earnings. Conversely, the marginal value of science education might be lower for individuals from such backgrounds, compared to individuals from socially and educationally disadvantaged groups. Thus, the value of science education by social and educational background is an open empirical question. We explore this question by estimating two equations, the first of which interacts *Science* with an indicator variable representing membership of a Scheduled Caste, and the second where *Science* is interacted with the parental education.

Panel D of Table 5 reports that significant and large returns to studying science for members of castes higher in the social hierarchy. Overall returns are 25% for individuals in the highest "General" category ($p < 0.01$) and 20% for the Other Backward Classes (in the middle of the social hierarchy ($p < 0.01$), but 15% and statistically indistinguishable from the null for the Scheduled Castes and Tribes.

In Panel E of the same table, the returns to science education are greatest for individuals with high household education (+26%, $p < 0.01$), followed by medium (+21%, $p < 0.01$) and low household education (+16%, $p < 0.01$).[22] This pattern holds when examining by ability subsamples specified earlier. Combined, these results point to the social environment as complementary to science education, with the greatest returns accruing to individuals who have social and parental support for translating their STEM skills into higher earnings.

[22] These are based on marginal effects calculated at the mean value of household average years of education of 10 (classified as medium education), and at values of average education one standard deviation higher (high education: 14 years) and one standard deviation lower (low education: 6 years).

4.3 Plausible channels

This section analyzes the role of two potential channels through which STEM education can lead to greater earnings. First, studying science in higher secondary grades might be associated with greater participation and completion of higher education, which would subsequently lead to increased incomes (Castello-Climent et al., 2018). Second, the combination of science in high school with more years of education might shift the sector (private or public) or type (salaried or business) where students are employed.

Table 6 estimates equation (1) using three different measures of educational attainment. Panel A of the table examines the result of studying science on the years of post-secondary education, Panel B reports whether the respondent at least completed a bachelor's degree (or equivalent), and Panel C whether the respondent completed any professional program (defined in the previous section). We find that science education at the secondary school stage is associated with 0.22 additional years of post-secondary education ($p < 0.01$). One potential explanation is selection into science, where motivation explains both the decision to study science as well as persistence within higher education. Alternatively, studying science could preserve more options for post-secondary education, which allows students to continue education more easily compared to non-science students. Corresponding to this finding, science students are also 5% more likely to complete a bachelors degree (Panel B, $p < 0.05$), and 6% more likely to complete a professional degree (Panel C, $p < 0.01$).

The labor market for educated men in urban India is classified into one of three types of employment: a position in the private sector, a relatively secure job in the public sector or running one's own business enterprise. Panels A and B in Table 7 show that studying science makes one more likely to get a public sector job, but only among low ability science students. However, we do not find effect on both private sector employment (Panel B) or business employment (Panel C). Thus, among lower ability students, science education makes one more likely to be in public sector relative to private sector. While public sector jobs are demanded by a relatively large section of society, high paying private jobs do compete in wages. But such private jobs often select those with better education. Thus for high ability students, studying science makes them equally likely to be in different kinds of employment. However, for low ability students, many good private jobs may not be available, leading them to prefer the public sector. For such students, technical backgrounds that prepare candidates better for selection examinations common for public sector positions may raise their chance of bagging a public sector job.

With this selection mechanism in the background, what implication does this have on incomes, conditional of being selected in a particular kind of job. Panels A, B and C in Table 8 report relatively high returns to science within each sector. The returns to science are the highest at almost 42% among high ability students who are businessmen. The return within the private sector, paradoxically, is not very different from the public sector; the point estimates in-fact suggest that the returns to science are slightly higher in the public sector jobs. But on the whole, studying science is equally useful whether one is in the public or the private sector. The story is slightly different for those with lower scholastic ability. Among low ability businessmen, having studied science gives no returns. This is equally true for the low ability students working in the public sector. This together with the result that science education raises the probability of being in public sector for low ability students, points out to the fact that a science education gets such individuals over the threshold of a government job but no further. There are returns to science however in the private sector, among low ability students.[23]

5 Robustness

Our regression model controls for ability by inclusion of dummy variable for divisions. However, there may be other variables: for example, other kinds of unobserved abilities not completely subsumed by scholastic performance, and other households factors that may potentially bias our results. In this section we assess the extent of potential omitted variable

[23]The public sector wages are however almost 89 percent higher than the private sector salaries for the low ability individuals, which is why they still seek public sector jobs.

bias due to such unobservable factors in the model. We investigate the extent of such omitted variable bias following a strategy developed by Altonji et al. (2005) and Oster (2017). This methodology is based on the idea that selection on observables can provide a useful guide to assess the selection based on unobservables. To elaborate further, let

$$Y = \beta_s X + \beta_z Z + W \tag{2}$$

where X is the main variable of interest, Z is observed and W contains all the unobserved components. The objective is to estimate the bias on β_1 because of W. The Altonji et al. (2005) methodology estimates this bias by positing an assumption:

$$\frac{Cov(X, W)}{Var(Z)} = \delta \frac{Cov(X, \beta_z Z)}{Var(\beta_z Z)} \tag{3}$$

In other words, the relation of X and unobservables is proportional to the relation of X to observables, the degree of proportionality given by δ. This basic insight has been extended by Oster (2017) to incorporate the idea that one can look at coefficient movements (of β_s) when covariates are added and deduce a similar bias. This extension also keeps account of movement in the R-squared value due to addition of control variables. Following this method, it is possible to derive a consistent estimator for the effect of *Science* as a function of two parameters: δ and R_{max}, denoted by $\beta_s(R_{max}, \delta)$. R_{max} is the R-square of a hypothetical regression which includes the complete set of controls including the unobservable variables. To operationalize this method, as a first step, we would need a baseline regression to which subsequent controls would be added. We posit a baseline regression where log of earnings is regressed on *Science*. As a second step, one needs to posit R_{max}. One way this could be set is by looking at R-squares obtained in other studies in the same context that control for the omitted variables. While there are mincerian returns to education regressions for India in the literature, there are none that look at the earnings of urban males who have passed high school.[24] Given the lack of a known R_{max}, we follow the other method suggested by Oster (2017) which sets R_{max} as 1.3 times the R-square of the regression that controls for Z (controlled regression). Since in our case, the R-square in our main specification is 0.304, we set R_{max} as 0.4. The robustness check suggested by Oster (2017) is that the interval $[\beta_s^{controlled}$, $\beta_s(\min(1.3{*}R^2_{controlled}, 1), 1)]$ should not contain 0. We find that this is indeed not the case (Table 9). In our case the $\beta_s(0.4, 1)]$ is 0.16. Moreover we also provide the value of δ for which β_s would become 0. The obtained value of 4 is high since Oster (2017) found that the average value of δ was 0.545 with 86% of the values of δ falling within $[0, 1]$. Alternatively we show the R_{max} that would needed to make β_s equal to zero, when δ equals 1. This value is 0.6, almost twice the R-square from the controlled regression. Thus, these robustness exercises suggest that the estimated returns to *Science* are robust.

6 Other Correlates of Science Choice

In previous sections, we have shown that there are positive returns to science education and these returns are robust: that is, as long as the relation between science and the known covariates allow us to learn about the relation between science choice and unknown covariates, tests point out the returns to science are considerable. In this section, we throw light on what some of these unknown characteristics could be: both cognitive as well as non-cognitive.

Data on these characteristics are hard to get at a national level especially in developing countries, since no large survey collects data on these dimensions. Hence we use a primary survey of high school students in two large states of India: Andhra Pradesh and Bihar, that were conducted by a team led by us. The choice of these states was purposive as they have a high proportion of science students in addition to representing a geographical north-south spread.[25] While we do not make any claims of representability for the nation or the state,

[24] Azam (2012) uses a sample of urban male wage earners to calculate returns to education. However, due to the nature of the dataset used in that paper, business employees are excluded. Moreover, the sample considered includes all adult males and not just those who have passed high school.

[25] In Bihar, 65 percent of students in the age group 16-18 and who attend school choose to study science. In Andhra Pradesh, this percentage is 91 percent. These are in contrast to 55 percent for the country as a

this is a representative survey of high school students in 6 cities in India. Two of the cities are large (Patna in Bihar and Vijayawada in Andhra Pradesh have a population above one million) while the other 4 (Bhagalpur and Sitamarhi in Bihar and Srikakulam and Kurnool in Andhra Pradesh) are mid-size cities. Apart from usual idiosyncratic differences, to the best of our knowledge, there is nothing remarkably different about these cities. The survey collected information on 558 students in class 12 (the last year of high-school) across 44 schools spread across the two states. Apart from information on subjects chosen by students and their life and career aspirations, the survey also measured various behavioral parameters: grit, ambiguity aversion, cognition-reflection ability and positive personality traits. We use this data to provide some suggestive evidence of the difference between students studying different high school majors in Table 10.

Individual behavioral traits, such as grit have been found to correlate with educational success and passion for long-term goals (Bowman et al., 2015; Duckworth et al., 2007). Its effect on these outcomes has been found to be present even after controlling for IQ and Big Five conscientiousness. Thus grit is likely to be an important factor in the labor market and hence to economic returns. Do they vary across those who choose science and those who do not? Using standardized questions suggested by the psychology literature (Duckworth et al., 2007), we score the sampled students on a grit scale (for more see Appendix 8).

On this trait, however, we find no difference between those choosing science and non-science majors. However, when it comes to positive personality, we find that science students score higher on this personality index as compared to non-science students. As explained in the Appendix 8, this index gives a higher score when students agree to a set of positive statements related to personality and non-cognitive skills (the higher the level of agreement to the statement, the higher the score on a five-point scale). This evidence is important because socio-emotional skills (personality traits and behaviors) have been emphasized in the recent literature (Heckman and Kautz (2012)) and have been found to be important in

whole.

many labor markets (Acosta et al. (2015), Deming (2017)).

An important correlate of subject choice is expected returns to such a choice in the future. However making such choices, weighing relative costs and benefits, requires rational decision making involving time discounting. This requires a certain level of cognitive ability. We use a three-item Cognitive Reflection Test (CRT), suggested by Frederick (2005), as a simple measure of one type of cognitive ability. This measure is predictive of the types of choices that are used to test expected utility theory and prospect theory. Using these tests, one finds that those who score higher on CRT are generally more patient. In our sample, Science students have a significantly higher CRT, pointing out that the choice of science may select students who are more willing to wait for long-run returns.

What is implicit in the capability to process information is also the assumption that students have information on labor market returns to jobs that are likely to follow and a reasonable idea of the probability of getting such jobs. However, such information may not be available to a high school student (or his/her parents). Our qualitative survey reveals that students and their parents have very poor knowledge of options that follow subject choice.[26] It is plausible that many students may not even want to make substantive choices till they have better information on the labor market. Recall that the choice of studying science leaves options open to study all subjects whereas studying business or humanities reflects a substantive decision: opting out of science-related professions, including the lucrative college degrees of engineering and medicine. It is then possible that the choice of a high school major is correlated to whether a student is willing to make decisions in ambiguous situations. Hence, a reasonable hypothesis is that those who make the substantive decision in high school: that chooses not to study science, are relative less ambiguous averse. We measure ambiguity aversion in two ways. We use an ambiguity tolerance scale suggested by the psychology literature (MSTAT-II) as well as ambiguity experiments suggested by Ellsberg (Appendix 8). Results show that while science and non-science students are equally ambiguity tolerant,

[26] Many students and parents were even unable to name their dream institutions post-school as well as jobs that would follow.

science students are significantly likely to pick a box that is less ambiguous (Box 1 where the exact number of red and blue balls are known) as compared to other boxes that represent more ambiguity. This is then consistent with the idea that taking a science major in school is correlated with the student's ambiguity aversion. Putting the results on CRT and ambiguity together, our evidence suggests that science students prefer to make choices later and they are patient enough to do so.

7 Conclusion

We explore the role of science education in grades 11 and 12, a stage where important career choices are made in Indian secondary education, on subsequent education, career and earnings outcomes. Our analysis, though not causal, shows that science education is associated with 21% higher earnings compared to humanities and business. We find that science education complements academic ability, English fluency, computer skills, parental education, and privileged social background, pointing to the importance of supporting these among disadvantaged students.

Globally, there is active debate over the value of STEM versus a traditional liberal arts education in today's digital economy because there are more jobs for students studying STEM – science, technology, engineering and math compared to a liberal arts major such as political science, philosophy or history. Policy makers in the U.S. have not only posed the choice between STEM and liberal arts education as a *substitute* but from President Obama on down, public officials have cautioned against pursuing degrees like art history, which are seen as expensive luxuries in today's world.[27] Results from this paper can provide useful insight to policy makers as we find evidence of strong complementarity. For example, computer and English fluency skills on top of a science degree enables workers to find higher paying jobs that have better career trajectory. Our results on complementarity is consistent

[27] On the contrary, when unveiling a new edition of the iPad, Steve Jobs explained that "it's in Apple's DNA that technology alone is not enough – that it's technology married with liberal arts, married with the humanities, that yields us the result that makes our hearts sing."

with Berman et al. (2003) and Lang and Siniver (2009) who find evidence of language-skill complementarity in the context of Israel. They show that improved Hebrew and English in addition to their native language accounts for 2/3 to 3/4 of the differential in earnings growth between immigrant and native employed in high-skilled occupation. Taken together, results from this paper implies the importance of complementary policies that can directly improve the ability of members of a disadvantaged group to undertake human capital investments.

Our paper also suggests that lack of information to students may be making some students postpone making irreversible choices. Studying science keeps all options open, leaving substantive choices for later when more information is revealed. This is consistent with our finding that science students are more ambiguity averse. That such students are also patient makes this a strategy to deal with an environment where little is known about consequences of choices. This calls for labor market information dissemination through a credible mechanism, thus allowing ambiguity averse individuals to make choices earlier in their careers.

Our results should be read with a number of caveats. First, in the absence of experimental or quasi-experimental research methodology, we cannot claim causality. Establishing causality of the effects of STEM education on professional outcomes might reveal the relative importance of selection versus treatment effects of science education, which is important for understanding the underlying production function as well as suggesting policy measures. A related issue is that we do not analyze potential barriers to students picking science, as well as the effectiveness of different pedagogical approaches to science education. We hope that these issues will be addressed in future research. ■

References

Acosta, P., N. Muller, and M. Sarzosa (2015). Beyond qualifications: Returns to cognitive and socio-emotional skills in Colombia.

Altonji, J., P. Arcidiacono, and A. Maurel (2015). The analysis of field choice in college and graduate school: Determinants and wage effects. NBER Working Paper 21655.

Altonji, J., E. Blom, and Meghir (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics 4*, 185–223.

Altonji, J., T. Elder, and C. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy 113*(1), 151–184.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics 121*, 343–375.

Azam, M. (2012). Changes in wage structure in urban India, 1983-2004: A quantile regression decomposition. *World Development 40*(6), 1135–1150.

Azam, M. (2016). Private tutoring: Evidence from India. *Review of Development Economics 20*(4), 739–761.

Azam, M., A. Chin, and N. Prakash (2013). The returns to English-language skills in India. *Economic Development and Cultural Change 61*(2), 335–367.

Berman, E., K. Lang, and E. Siniver (2003). Language-skill complementarity: Returns to immigrant language acquisition. *Labour Economics 10*, 265–90.

Black, S., Z. He, C. Muller, and A. Spitz-Oener (2015). On the origins of STEM: The role of high school STEM coursework in occupational determination and labour market success in mid life.

Bowman, N., P. Hill, N. Denson, and R. Bronkema (2015). Keep on truckin' or stay the course? Exploring grit dimensions as differential predictors of educational achievement, satisfaction, and intentions. *Social Psychological and Personality Science 6*(6), 639–645.

Brown, C. and M. Corcoron (1997). Sex based differences in school content and the male-female wage gap. *Quarterly Journal of Economics 99*, 31–44.

Card, D. (1999). The causal effects of education on earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A, Chapter 30, pp. 1801–1863. Elsevier Science B.V.

Castello-Climent, A., L. Chaudhary, and A. Mukhopadhyay (2018). Higher education and prosperity: From Catholic missionaries to luminosity in India. *Economic Journal*.

Chakraborty, T. and S. Kapur (2009). English language premium: Evidence from a policy experiment in India. *Economics of Education Review 50*, 1 – 16.

Daymont, T. and P. Andrisani (1984). Job preferences, college major, and the gender gap in earnings. *Journal of Human Resources 19*, 408–428.

Deming, D. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*.

Duckworth, A., C. Peterson, M. Matthews, and D. Kelly (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology 92*(6), 1087–1101.

Duraiswamy, P. (2002). Changes in returns to education in India, 1983-94: By gender, age-cohort and location. *Economics of Education Review 21*(6), 609–622.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives 19*(4), 1087–1101.

Gemici, A. and M. Wiswall (2014). Evolution of gender differences in post-secondary human capital investments: College majors at the intensive margins. *International Economic Review 55*, 23–56.

Grogger, J. and E. Eide (1995). Changes in college skills and the rise in the college wage premiums. *Journal of Human Resources 30*, 280–310.

Hastings, J., C. Neilson, and S. Zimmerman (2013). Are some degrees worth more than others? Evidence from college admissions cutoffs in Chile. NBER Working Paper 19241.

Heckman, J. and T. Kautz (2012). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464.

Jain, T., P. Maitra, and S. Mani (2016). Barriers to skill acquisition: Evidence from English training in India. IZA Discussion Paper No. 10199.

James, E., A. Nabeel, J. Conaty, and D. To (1989). College quality and future earnings: Where should you send your child to college? *American Economic Review 79*, 247–252.

Kirkeboen, L., E. Leuven, and M. Mogstad (2015). Field of study, earnings, and self-selection. University of Chicago Working Paper.

Klasen, S. and Pieters, J. (2015). What Explains the Stagnation of Female Labor Force Participation in Urban India? *The World Bank Economic Review 29*(3).

Lang, K. and E. Siniver (2009). The return to English in a non-English speaking country: Russian immigrants and native Israelis in Israel. *The B.E. Journal of Economic Analysis & Policy 9*, 1–30.

Loury, L. (1997). The gender-earnings gap among college-educated workers. *Industrial and Labor Relations Review 50*, 580–593.

Loury, L. and D. Garman (1995). College selectivity and earnings. *Journal of Labor Economics 13*, 289–308.

McLain, D. (2009). Evidence of the properties of an ambiguity tolerance measure: The multiple stimulus types ambiguity tolerance scale–II (MSTAT–II). *Psychological Reports 105*(3), 975–988.

MHRD, Government of India (1998). National policy on education 1996 (modified in 1992).

Munshi, K. and M. Rosenzweig (2006). Traditional institutions meet the modern world: Caste, gender and schooling choice in a globalizing economy. *American Economic Review 96*(4), 1225–1252.

OECD (2015, April). Education Indicators in Focus 31.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 0*(0), 1–18.

Weinberger, C. (1998). Race and gender wage gaps in the market for recent college graduates. *Industrial Relations 37*, 67–84.

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

World Economic Forum (2016). The Human Capital Report 2016.

Table 1: Summary Statistics

	Mean	Standard Deviation
Dependent Variables:		
Log(Earnings)	4.75	0.99
Annual Earnings (Rs. '000s)	178.33	212.23
Years of Education (Post Grade 10)	3.86	1.84
Dummy: At least Graduate Education	0.57	0.49
Dummy: Professional Education	0.06	0.23
Dummy: Private Employment	0.25	0.43
Dummy: Public Employment	0.26	0.44
Dummy: Business Employment	0.27	0.44
Independent Variables:		
Science Major	0.25	0.43
Business Major	0.23	0.42
Humanities Major	0.52	0.5
Division I	0.32	0.47
Division II	0.57	0.5
Division III	0.12	0.32
Repeated Grade	0.12	0.32
Fluent English	0.36	0.48
Less Fluent English	0.48	0.5
Demographic Controls:		
Age	39.81	10.2
Married	0.83	0.38
Scheduled Castes	0.12	0.32
Scheduled Tribes	0.03	0.17
Other Backward Class	0.33	0.47
Muslim	0.08	0.27
Christian	0.03	0.17
Average Household Education	10.02	3.84
Max Parent Education	8.26	4.96
Observations	4763	

NOTES: Mean and standard deviation of the estimation sample is reported. The number of observations for the variables Average Household Education and Max Parent Education are 4,687 and 2,513 respectively.

Table 2: Returns to High School Science Major

Dependent Variable:	Log(Earnings)				
	No Control	Ability FE	District FE	Demographics	Parent Edu
	(1)	(2)	(3)	(4)	(5)
Science	0.36***	0.20***	0.25***	0.22***	0.21***
	(0.06)	(0.05)	(0.04)	(0.04)	(0.05)
Ability Controls:					
Dummy: 1st Division		0.33***	0.22***	0.21***	0.18*
		(0.08)	(0.07)	(0.07)	(0.10)
Dummy: 2nd Division		0.08	0.02	0.02	0.00
		(0.07)	(0.06)	(0.06)	(0.08)
Dummy: Repeated Grade		-0.32***	-0.25***	-0.22***	-0.25***
		(0.05)	(0.05)	(0.05)	(0.06)
Dummy: Less Fluent English		0.08	0.14**	0.11**	0.08
		(0.09)	(0.05)	(0.04)	(0.05)
Dummy: Fluent English		0.41***	0.42***	0.35***	0.36***
		(0.11)	(0.06)	(0.05)	(0.06)
Demographic Controls:					
Age				0.06***	0.03
				(0.02)	(0.02)
Age Square				-0.00***	-0.00
				(0.00)	(0.00)
Dummy: Married				0.09	0.07
				(0.07)	(0.06)
Dummy: Scheduled Castes				-0.06	-0.06
				(0.04)	(0.07)
Dummy: Scheduled Tribes				0.02	-0.25
				(0.12)	(0.18)
Dummy: Other Backward Class				-0.03	-0.01
				(0.04)	(0.07)
Dummy: Muslim				-0.01	0.07
				(0.06)	(0.12)
Dummy: Christian				0.07	0.16**
				(0.06)	(0.07)
Average Household Education				0.03***	0.03***
				(0.00)	(0.01)
Max Parent Education					0.01
					(0.01)
Constant	4.65***	4.39***	4.41***	2.58***	3.07***
	(0.06)	(0.13)	(0.08)	(0.30)	(0.32)
Observations	4,763	4,763	4,763	4,687	2,513
R-squared	0.03	0.10	0.25	0.30	0.36

NOTES: Robust standard errors clustered at state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Returns to High School Science Major by Quintiles

Dependent Variable:	Log(Earnings)					
	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}	99^{th}
	(1)	(2)	(3)	(4)	(5)	(6)
Science	0.25***	0.18***	0.20***	0.23***	0.25***	0.37***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Constant	3.07	3.79	3.79*	4.49***	4.73	6.07
	(3.36)	(5.21)	(1.95)	(0.25)	(4.68)	(10.06)
Observations	4,687	4,687	4,687	4,687	4,687	4,687

NOTES: This table reports the coefficients corresponding to specification (4) of Table 2. All specifications control for ability, demographics and district fixed effects. Each cell is the estimated coefficient of choosing Science major from separate regressions. Robust standard errors clustered at state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Returns to High School Science Major by Ability

Dependent Variable:	Log(Earnings)	
	1st Division	2nd/3rd Division
	(1)	(2)
Science	0.25***	0.19***
	(0.07)	(0.06)
Constant	3.54***	2.35***
	(0.73)	(0.33)
Observations	1,497	3,190
R-squared	0.36	0.30

NOTES: This table reports the coefficients corresponding to specification (4) of Table 2. All specifications control for ability, demographics and district fixed effects. Each column is the estimated coefficient of choosing Science major from separate regressions. Robust standard errors clustered at state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Heterogeneity in Returns to High School Science Major

	Full Sample	Division I	Division II/III
	(1)	(2)	(3)
PANEL A: Language Proficiency			
Fluent English	0.28***	0.28***	0.27***
	(0.05)	(0.10)	(0.09)
Little English	0.19***	0.23**	0.21***
	(0.05)	(0.11)	(0.04)
No English	0.13	0.18	-0.06
	(0.10)	(0.14)	(0.17)
PANEL B: Computer Proficiency			
Computer: Yes	0.26***	0.31***	0.19**
	(0.05)	(0.08)	(0.09)
Computer: No	0.07*	-0.03	0.11**
	(0.04)	(0.09)	(0.05)
PANEL C: Professional Degree			
Professional Edu: Yes	0.20	0.12	0.37
	(0.13)	(0.19)	(0.24)
Professional Edu: No	0.20***	0.23***	0.17***
	(0.04)	(0.08)	(0.05)
PANEL D: Caste Groups			
Caste Group: General	0.25***	0.32***	0.21**
	(0.05)	(0.07)	(0.09)
Caste Group: OBC	0.20***	0.19*	0.16***
	(0.05)	(0.11)	(0.05)
Caste Group: SC/ST	0.15	0.12	0.21**
	(0.09)	(0.16)	(0.10)
PANEL E: Household Education			
Household Edu: High	0.26***	0.32***	0.21**
	(0.05)	(0.07)	(0.08)
Household Edu: Medium	0.21***	0.22***	0.19***
	(0.04)	(0.08)	(0.06)
Household Edu: Low	0.16***	0.12	0.17**
	(0.04)	(0.10)	(0.08)

NOTES: This table reports the marginal effects of studying Science. All specifications control for ability, demographics and district fixed effects. Column 1 reports the marginal effect by various indicators: Language Proficiency, Computer Proficiency, Professional Degree, Caste Groups and Household Education. Column 2 and 3 report the similar marginal effects by divisions (I and II & III). Each panel is a separate regression. Robust standard errors clustered at state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Returns to High School Science Major: Human Capital Outcomes

	Full Sample	Division I	Division II/III
	(1)	(2)	(3)
Dependent Variable:	PANEL A: Years of Education		
Science	0.22***	0.25**	0.23***
	(0.07)	(0.11)	(0.07)
Constant	1.98***	3.68***	1.58***
	(0.51)	(0.75)	(0.53)
R-squared	0.33	0.36	0.30
Dependent Variable:	PANEL B: Graduate Education		
Science	0.05**	0.06*	0.06***
	(0.02)	(0.03)	(0.02)
Constant	0.10	0.55*	0.03
	(0.16)	(0.30)	(0.13)
R-squared	0.30	0.33	0.27
Dependent Variable:	PANEL C: Professional Education		
Science	0.06***	0.08***	0.04**
	(0.01)	(0.01)	(0.02)
Constant	0.13	0.24	0.07
	(0.08)	(0.14)	(0.07)
R-squared	0.13	0.22	0.11
Observations	4,687	1,497	3,190

NOTES: This table reports the coefficients corresponding to specification (4) of Table 2. All specifications control for ability, demographics and district fixed effects. Each column is the estimated coefficient of choosing Science major from separate regressions by divisions (Full Sample, I and II & III). Robust standard errors clustered at state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Returns to High School Science Major: Employment Outcomes

	Full Sample	Division I	Division II/III
	(1)	(2)	(3)
Dependent Variable:	PANEL A: Public Employment		
Science	0.02	-0.00	0.04*
	(0.02)	(0.04)	(0.02)
Constant	-0.73***	-0.57	-0.67***
	(0.15)	(0.37)	(0.17)
R-squared	0.21	0.29	0.22
Dependent Variable:	PANEL B: Private Employment		
Science	-0.02	-0.01	-0.02
	(0.02)	(0.04)	(0.02)
Constant	0.55***	0.59**	0.56***
	(0.12)	(0.27)	(0.12)
R-squared	0.14	0.22	0.17
Dependent Variable:	PANEL C: Business Employment		
Science	0.01	-0.02	0.02
	(0.02)	(0.03)	(0.02)
Constant	0.39***	0.15	0.43***
	(0.11)	(0.22)	(0.12)
R-squared	0.17	0.24	0.20
Observations	4,687	1,497	3,190

NOTES: This table reports the coefficients corresponding to specification (4) of Table 2. All specifications control for ability, demographics and district fixed effects. Each column is the estimated coefficient of choosing Science major from separate regressions by divisions (Full Sample, I and II & III). Robust standard errors clustered at state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Returns to High School Science Major: Income from Employment Outcomes

	Full Sample	Division I	Division II/III
	(1)	(2)	(3)
Dependent Variable:	PANEL A: Income from Public Employment		
Science	0.18***	0.26***	0.13
	(0.05)	(0.07)	(0.08)
Constant	3.60***	5.01***	2.72***
	(0.64)	(1.38)	(0.50)
R-squared	0.48	0.52	0.56
Observations	1,209	488	721
Dependent Variable:	PANEL B: Income from Private Employment		
Science	0.23***	0.21*	0.21***
	(0.05)	(0.10)	(0.07)
Constant	3.66***	4.32***	3.36***
	(0.40)	(0.64)	(0.41)
R-squared	0.46	0.50	0.48
Observations	1,167	400	767
Dependent Variable:	PANEL C: Income from Business Employment		
Science	0.18*	0.42*	0.08
	(0.10)	(0.22)	(0.11)
Constant	2.36***	2.74	2.10***
R-squared	0.37	0.53	0.40
Observations	1,273	320	953

NOTES: This table reports the coefficients corresponding to specification (4) of Table 2. All specifications control for ability, demographics and district fixed effects. The sample in Panel A, B and C consists of all public salaried individuals, all private salaried individuals and individuals employed in business respectively. Each column is the estimated coefficient of choosing Science major from separate regressions by divisions (Full Sample, I and II & III). Robust standard errors clustered at state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Robustness to Omitted Variable Bias

	Coefficient of Science				
	Uncontrolled	*Controlled*	Identified (Estimated Bias)		
			$R^2_{max} = 0.4$		$\delta = 1$
			β_s for $\delta = 1$	δ for $\beta = 0$	R^2_{max} for $\beta = 0$
β_s	0.36	0.22	0.16	3	0.6
R^2	0.03	0.30			

NOTES: We follow Oster (2017) to formally test for robustness to omitted variable bias by observing the coefficient movements after inclusion of controls. $R^2_{max} = 1.3 * R^2_{controlled} = 0.4$. This is based on recommendations made in Oster (2017).

Table 10: Science versus Non-Science Major Choice

	Non-Science		Science		Difference
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	
Grit score	219	3.369	339	3.460	-0.091
	[38]	[0.041]	[44]	[0.053]	
Ambiguity score	219	40.192	339	40.339	-0.147
	[38]	[0.477]	[44]	[0.566]	
Ambiguity Experiment Box 1	219	0.393	339	0.472	-0.079*
	[38]	[0.035]	[44]	[0.027]	
Ambiguity Experiment Box 1 and 2	219	0.562	339	0.614	-0.052
	[38]	[0.040]	[44]	[0.026]	
Ambiguity Experiment Box 1, 2 and 3	219	0.772	339	0.791	-0.019
	[38]	[0.031]	[44]	[0.017]	
CRT Score	122	0.574	230	0.874	-0.300***
	[24]	[0.099]	[28]	[0.096]	
Personality Score	219	30.918	339	31.676	-0.758*
	[38]	[0.361]	[44]	[0.216]	
F-test of joint significance (F-stat)					21.222***
F-test, number of observations					209

NOTES: The value displayed for t-tests are the differences in the means between science and non-science groups. The value displayed for F-tests are the F-statistics. Standard errors are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

Table 11: Returns to High School Science Major (including Unemployed People)

Dependent Variable:	Log(Earnings)				
	No Control (1)	Ability FE (2)	District FE (3)	Demographics (4)	Parent Edu (5)
Science	0.23***	0.09	0.18***	0.17***	0.15**
	(0.08)	(0.06)	(0.05)	(0.04)	(0.06)
Ability Controls:					
Dummy: 1st Division		0.39***	0.22**	0.24**	0.24**
		(0.10)	(0.09)	(0.10)	(0.11)
Dummy: 2nd Division		0.18**	0.09	0.09	0.09
		(0.09)	(0.08)	(0.08)	(0.10)
Dummy: Repeated Grade		-0.36***	-0.31***	-0.25***	-0.29***
		(0.07)	(0.06)	(0.06)	(0.07)
Dummy: Fluent English		0.30**	0.34***	0.30***	0.25**
		(0.12)	(0.06)	(0.06)	(0.10)
Dummy: Less Fluent English		-0.00	0.05	0.04	-0.02
		(0.11)	(0.04)	(0.04)	(0.08)
Demographic Controls:					
Age				0.13***	0.11***
				(0.02)	(0.03)
Age Square				-0.00***	-0.00**
				(0.00)	(0.00)
Dummy: Married				0.54***	0.46***
				(0.09)	(0.08)
Dummy: Scheduled Castes				-0.10*	-0.22***
				(0.05)	(0.08)
Dummy: Scheduled Tribes				0.03	-0.51**
				(0.13)	(0.25)
Dummy: Other Backward Class				0.01	0.03
				(0.04)	(0.09)
Dummy: Muslim				0.04	0.12
				(0.08)	(0.14)
Dummy: Christian				0.11	0.21
				(0.13)	(0.16)
HH Education				0.02***	0.03**
				(0.01)	(0.01)
Max Parent Education					-0.01
					(0.01)
Constant	4.47***	4.22***	4.25***	0.54	1.05*
	(0.08)	(0.17)	(0.08)	(0.51)	(0.59)
Observations	5,001	5,001	5,001	4,925	2,737
R-squared	0.01	0.04	0.17	0.28	0.32

NOTES: Robust standard errors clustered at state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 12: Summary Statistics from Survey

	No of Obs.	Mean	Std. dev.
Age of the student	880	16.885	0.94
Math score in class X	619	70.864	18.83
Science score in class X	605	69.107	16.79
CRT score of the student	461	0.681	0.91
English Score in class X	525	71.598	19.27
Grit score of the student	884	3.417	0.61
Mother completed class X	884	0.673	0.47
Father completed class X	884	0.834	0.37
Household Size	884	4.827	1.96
Distance to closest bank (in kms.)	883	2.567	6.03
Religion: Hindu	884	0.880	0.33
Religion: Muslim	884	0.068	0.25
Female Student	884	0.408	0.49
Personality score of the student	884	31.189	3.26
Ambiguity score	884	39.991	9.05
Scheduled Caste	884	0.197	0.40
General Caste	884	0.279	0.45
Other Backward Caste	884	0.508	0.50
Tier I city	884	0.393	0.49
Tier II city	884	0.360	0.48
Tier III city	884	0.248	0.43
CBSE Syllabus in class X	884	0.213	0.41
ICSE Syllabus in class X	884	0.110	0.31
State Syllabus in class X	884	0.672	0.47

NOTES: Data is obtained from the Career Choice Primary Survey the authors conducted in 2017. The table reports mean and standard deviations of the variables used in subsequent regressions.

Table 13: Summary Statistics from Survey

	No of Obs.	Mean	Std. dev.
Electric Connection	884	0.988	0.11
Land line telephone	884	0.037	0.19
Internet connection	884	0.328	0.47
Tap Water supply	884	0.695	0.46
Does the student have access to cell phone?	883	0.535	0.50
Does his/her phone have internet access?	871	0.359	0.48
Friend took science	884	0.689	0.46
Friend took commerce	884	0.238	0.43
Friends took arts	884	0.118	0.32
First Division	884	0.612	0.49
Second Division	884	0.136	0.34
Third Division	884	0.126	0.33
Bihar	884	0.521	0.50
Parent gave a lot of thought on student's education	884	0.483	0.50
Parent thinks that stream choice is an important signal	884	0.579	0.49
Parent thinks that stream choice is important for job	884	0.523	0.50
Student gave a lot of thought on his/her stream choice	884	0.770	0.42
Student thinks that science stream is for smarter students	884	0.262	0.44
Challenging career is important for student	884	0.764	0.43
Earnings is important for student	884	0.835	0.37
Career with travel opportunities is important for student	884	0.610	0.49
Career that allows to stay in a big city is important for student	884	0.689	0.46
Career that emphasizes managerial skills is important for student	884	0.569	0.50
Career that has non-transferable job is important for student	884	0.542	0.50
Referred to siblings for information	884	0.428	0.50
Referred to friends for information	884	0.249	0.43

NOTES: Data is obtained from the Career Choice Primary Survey the authors conducted in 2017. The table reports mean and standard deviations of the variables used in subsequent regressions.

Table 14: Correlates of Science Major Choice in High School

Dependent Variable:	Science Major					
	(1)	(2)	(3)	(4)	(5)	(6)
Math Score in Class X	0.0063***	0.0066***				
	(0.00)	(0.00)				
Science Score in class X	0.0061***	0.0065***				
	(0.00)	(0.00)				
Second Division			-0.25***	-0.22***		
			(0.05)	(0.05)		
Third Division			-0.56***	-0.50***		
			(0.06)	(0.08)		
Friend took Science					0.24***	0.16***
					(0.06)	(0.05)
Friend took Commerce					-0.15***	-0.18***
					(0.06)	(0.05)
Friend took Arts					-0.13**	-0.11*
					(0.06)	(0.05)
Household Controls	No	Yes	No	Yes	No	Yes
Demographic Controls	No	Yes	No	Yes	No	Yes
Other Controls	No	Yes	No	Yes	No	Yes
N	600	588	772	757	884	866
R2	0.19	0.23	0.17	0.22	0.12	0.20

NOTES: Data is obtained from the Career Choice Primary Survey the authors conducted in 2017. Dependent variable is a binary variable which is 1 if the student chose science in class XII. Errors are clustered at the school level. Household level controls include household size, father's education, mother's education, whether father is a salaried employee, whether mother is a salaried employee, asset index and distance to closest bank. Demographic controls include age, gender of student, caste and religion. Other controls such as class X state board syllabus, city, and state are also specified in the regression.

Table 15: Correlates of Science Major Choice in High School

Dependent Variable:	Science Major			
	(1)	(2)	(3)	(4)
Parent gave a lot of thought on student's education	0.11**	0.059		
	(0.05)	(0.05)		
Parent thinks that stream choice is important for job	0.045	0.057		
	(0.05)	(0.05)		
Student gave a lot of thought on his/her stream choice			0.17***	0.12***
			(0.04)	(0.04)
Student thinks that science stream is for smarter students			0.072*	0.13***
			(0.04)	(0.03)
Household Controls	No	Yes	No	Yes
Demographic Controls	No	Yes	No	Yes
Other Controls	No	Yes	No	Yes
N	884	866	884	866
R2	0.021	0.13	0.029	0.15

NOTES: Data is obtained from the Career Choice Primary Survey the authors conducted in 2017. Dependent variable is a binary variable which is 1 if the student chose science in class XII. Errors are clustered at the school level. Household level controls include household size, father's education, mother's education, whether father is a salaried employee, whether mother is a salaried employee, asset index and distance to closest bank. Demographic controls include age, gender of student, caste and religion. Other controls such as class X state board syllabus, city, and state are also specified in the regression.

Table 16: Correlates of Science Major Choice in High School

Dependent Variable:	Science Major			
	(1)	(2)	(3)	(4)
Challenging career is important for student	0.13**	0.099*		
	(0.06)	(0.05)		
Earnings is important for student	0.098*	0.054		
	(0.05)	(0.05)		
Career with travel opportunities is important for student	0.099**	0.080**		
	(0.04)	(0.04)		
Career that has non-transferable job is important for student	-0.070*	-0.060		
	(0.04)	(0.04)		
Career that allows to stay in a big city is important for student	-0.023	-0.024		
	(0.04)	(0.04)		
Career that emphasizes managerial skills is important for student	-0.14***	-0.14***		
	(0.05)	(0.05)		
Referred to elder siblings for information			-0.033	-0.020
			(0.03)	(0.03)
Referred to friends for information			-0.13***	-0.077
			(0.05)	(0.05)
Referred to mother for information			0.011	-0.015
			(0.05)	(0.04)
Referred to father for information			0.081*	-0.0090
			(0.04)	(0.04)
Referred to counsellor for information			0.026	0.0029
			(0.09)	(0.08)
Referred to teacher for information			0.034	0.042
			(0.04)	(0.04)
Household Controls	No	Yes	No	Yes
Demographic Controls	No	Yes	No	Yes
Other Controls	No	Yes	No	Yes
N	884	866	884	866
R2	0.061	0.17	0.025	0.13

NOTES: Data is obtained from the Career Choice Primary Survey the authors conducted in 2017. Dependent variable is a binary variable which is 1 if the student chose science in class XII. Errors are clustered at the school level. Household level controls include household size, father's education, mother's education, whether father is a salaried employee, whether mother is a salaried employee, asset index and distance to closest bank. Demographic controls include age, gender of student, caste and religion. Other controls such as class X state board syllabus, city, and state are also specified in the regression.

Table 17: Correlates of Science Major Choice in High School

Dependent Variable:	Science Major									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Grit Score	0.046	0.058**								
	(0.03)	(0.02)								
Ambiguity Tolerance Score			0.0018	0.0016						
			(0.00)	(0.00)						
Ambiguity Tolerance - Experiment					0.0052	-0.0026				
					(0.03)	(0.03)				
CRT Score							0.099***	0.058***		
							(0.03)	(0.02)		
Personality Score									0.015**	0.013**
									(0.01)	(0.01)
Household Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Other Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	884	866	884	866	884	866	461	453	884	866
R2	0.0082	0.13	0.0060	0.13	0.0049	0.13	0.035	0.18	0.014	0.13

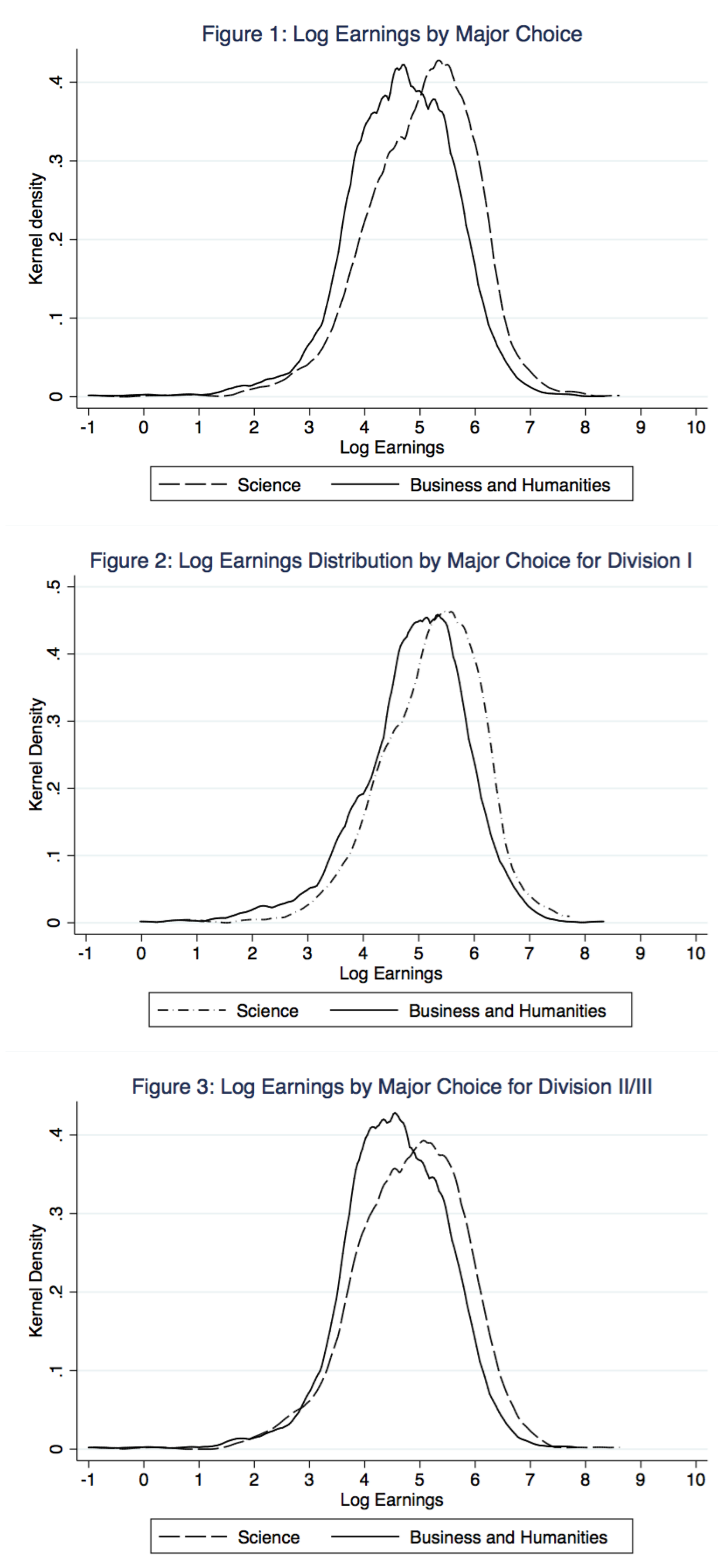
NOTES: Data is obtained from the Career Choice Primary Survey the authors conducted in 2017. Dependent variable is a binary variable which is 1 if the student chose science in class XII. Errors are clustered at the school level. Household level controls include household size, father's education, mother's education, whether father is a salaried employee, whether mother is a salaried employee, asset index and distance to closest bank. Demographic controls include age, gender of student, caste and religion. Other controls such as class X state board syllabus, city, and state are also specified in the regression.

Table 18: Factors Loading into Science - Using LASSO

Variable	Coefficient
Age of the student	-0.0717
Household Size	-0.0013
Mother Completed Class X	0.0537
Father is a salaried employee	0.0819
Parent gave a lot of thought on student's stream choice	0.0455
Parent thinks that stream choice is important for job	0.0125
State Syllabus in class X	-0.0285
Student thinks that science stream is for smarter students	0.0634
Student gave a lot of thought on his/her stream choice	0.0936
Ambiguity Tolerance - Experiment	-0.007
Challenging career is important for student	0.0662
Earnings is important for student	0.039
Career with travel opportunities is important for student	0.0508
Career that has non-transferable job is important for student	-0.0242
Career that allows to stay in a big city is important for student	-0.0226
Career that emphasizes managerial skills is important for student	-0.1104
Number of rooms in the household	0.0073
Friend choose science	0.2627
Referred to siblings for information	-0.0136
Referred to friends for information	-0.0369
Personality Score	0.0047

Variables are selected using the LASSO (Belloni et al., 2014, Bradley et al., 2004) that are predictive of the main outcome variable in our analysis. The table lists the complete set of variables which have been selected using the LASSO Method. The coefficients mentioned in the table are for the Least Angle Regression Model with the lowest Mallow's C_p statistic or the model with lowest mean squared prediction error.

Figure 1: Log(Earnings) Distribution by High School Science Major Choice



NOTES:

8 Appendix: Description of Variables

Variable Name	Variable Description
Grit	Grit is defined as the perseverance and passion for long term goals. We employ the 12-item Grit Scale developed by Duckworth et al. (2007). During the survey, students were asked to rate their agreeableness with each of the statements (items) in the grit scale according to a 5 point rating with 1 corresponding to 'Very much like me' and 5 corresponding to 'Not like me at all'. Student with a high score on the aggregated Grit Scale indicates a person with higher grit. Extant research has found that grit is positively associated with educational achievement, GPA scores and probability of completing a task which are important determinants of a successful career.
Ambiguity Score	Ambiguity Intolerance of the students were measured using the Multiple Stimulus Types Ambiguity Tolerance Scale - II (MSTAT - II). This 13-item psychometric scale developed by McLain (2009) assesses the cognitive response of participants to different ambiguous stimuli. Individual items were measured on a 5 point rating with 1 corresponding to 'Do not agree' and 5 corresponding to 'Completely agree'. Low scores on the Ambiguity Tolerance Scale indicate ambiguity intolerance and high scores indicate a liking for ambiguity.

Ambiguity Experiment	Students were presented with four boxes consisting of 10 blue and red balls in varying proportions. Box 1 contained 5 red and 5 blue balls. The second, third and fourth boxes contained anywhere between 4 and 6, 2 and 8, 0 and 10 blue balls respectively. They were then asked to pick a box from which a ball will be drawn at random. They win the game if the ball drawn is blue in color. From our data, we construct binary variables by combining Box 1, Box 1 and Box 2, and Box 1, Box 2 and Box 3 which equals various thresholds of ambiguity aversion. The game was an adaptation of the famous "Ellsberg Paradox" in which participants were found to prefer situations with known probabilities of events to situations where the probabilities of the events are unknown.

<table>
<tr>
<td>CRT Score</td>
<td>Cognitive Reflection Test (CRT) is a test of how quickly a student processes and responds to a basic aptitude question. The question didn't require any written calculation. This was to test whether the student responds with a obvious looking incorrect answer or processes the subtlety of the question thoroughly and responds with a correct answer. Each correct answer was awarded a point of 1 and the total scores for each student was calculated out of 3. Following are the questions:

• A bat and a ball cost Rs 110 in total. The bat costs Rs 100 more than the ball. How much does the ball cost?

• If it takes 5 machines 5 minutes to make 5 phones, how long would it take 100 machines to make 100 phones?

• In a closed container, there is an insect. Every day, the number of insects doubles. If it takes 48 days to fill the container. When was the container half filled?</td>
</tr>
</table>

Personality	This was to test the personality and non cognitive skills of the students. The students were presented with a set of positive statements related to personality and non cognitive skills and were asked to rate their agreeableness on a 5 point scale with 1 corresponding to 'Do not agree' and 5 corresponding to 'Completely agree'. These scores were aggregated for all statements for a student to get a cumulative personality score. Following are the statements: • I like to be very good at what I do. • I feel I can do just about anything if I put my mind to it. • I can be very disciplined and push myself. • I am often in a good mood. • I want to achieve more than my parents have • I am looking forward to a successful career. • I have high goals and expectations for myself.