I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11908

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

Tarun Jain
Abhiroop Mukhopadhyay
Nishith Prakash
Raghav Rakesh

OCTOBER 2018

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11908

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

Tarun Jain
Indian School of Business

Abhiroop Mukhopadhyay
Indian Statistical Institute and IZA

Nishith Prakash
University of Connecticut, IZA, HiCN and CReAM

Raghav Rakesh
Michigan State University

OCTOBER 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

This paper explores the association between studying science at the higher secondary stage and labor market earnings using nationally representative data on high school subject choices and adult outcomes for urban males in India. Results show that those who studied science in high school have 22% greater earnings than those who studied business and humanities, even after controlling for several measures of ability. These higher earnings among science students are further enhanced if the students also have some fluency in English. Moreover, greater earnings are observed among individuals with social and parental support for translating science skills into higher earnings. Science education is also associated with more years of education, likelihood of completing a professional degree, and among low ability students, working in public sector positions.

JEL Classification: I23, I26, J24

Keywords: high-school majors, labor markets, science, STEM, India

Corresponding author:
Tarun Jain
Indian School of Business
Hyderabad 500032
India
E-mail: tj9d@virginia.edu

* We appreciate helpful comments and suggestions from Mehtabul Azam, Eric Edmonds, Caroline Hoxby, Karthik Muralidharan, Stephen Ross, Duncan Thomas, Sarah Turner and numerous seminar and conference audiences. 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. Approved by IRBs at Indian School of Business and University of Connecticut and ethics committee at Indian Statistical Institute.

"[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 (OECD (2015), national statistics websites for China and India). Despite the success of several countries in producing STEM graduates, and the attempt of others to follow, the labor market consequences of STEM education are poorly understood. Understanding the influence of STEM on eventual career choices and earnings as well as the pathways that enable these outcomes is an important question for both researchers and policymakers, potentially helping to design policies that encourage students and administrators to pursue the most productive educational paths.

Given the country's success in training STEM graduates, by itself India is an important setting to study the influence of STEM. Our strategy for studying STEM is to first examine science education as a pathway into further STEM training and careers. For this, we use the Indian education structure's specialization in either science, business or humanities at the higher secondary stage, and then estimate the effect of science education on earnings and career choices.[2,3] Specifically, using nationally representative data from India, we show the influence of science education on subsequent post-secondary education, adult earnings

[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] A large share of students opt for science at this stage – among adults between age 25 and 65 years in India, approximately 22.3% had studied science, 16.6% had studied business and 61.1% had a background in humanities.

[3] In contrast, education systems in North America and the United Kingdom typically do not allow such specific focus.

and type of job, while illustrating the role of complementary factors such as English and computer skills, parental and social background.

Estimating the labor market consequences of science education 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 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 mitigate these biases, we exploit the richness of the India Human Development Survey (IHDS) that allows us to control for ability using performance in the high stakes 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. These controls allow us to report precise correlations between major choice and labor market earnings.

Concerns may still remain on whether our strategy provides causal estimates. A credible identification strategy would exogenously vary either the demand for or the supply of science education. However, such variations are not easy to come by.[4,5] Only experimental variation where students are randomly assigned to high school majors will overcome these identification challenges, something that is challenging in this context. We therefore take the route suggested by Altonji et al. (2005) and Oster (2017) and conduct bounds analysis on the estimated coefficients.

An alternate lens on our results is that we provide estimates that may be useful to

[4]The problems with causal interpretation of returns to human capital from non experimental data is not restricted only to developing countries. Recent studies in the context of the U.S., such as 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.

[5]It is clear that the exogenous variation in the supply of science education is a more promising strategy for identification since any variation in the demand for science education would involve differential perceptions of the returns to science education.

understand why India produces so many STEM graduates. Do science students, who subsequently feed into STEM streams, indeed earn more, even when compared to observationally equivalent students from other streams? Our estimates are also useful in providing a value to science education. Insofar as most returns to human capital estimates, in particular those estimated in developing countries, have similar problems in causal interpretation, our benchmark "returns" may provide interesting comparisons with other indicators of human capital in the literature: for example, to Mincerian returns, studying English, or knowing computers.[6]

We report several interesting results. First, studying science yields sizable earnings increase 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 for men who study science in high school relative to men who study business and humanities. Even after controlling for parental education, we find earnings are 21% higher for men who study science. Results from quantile regressions show that earnings are similar at all points of the wage distribution except the highest one percent – earnings associated with studying science are as high as 37% for income earners in the 99^{th} percentile.

Second, heterogeneity analysis by ability suggest differential earnings from studying science relative to business and humanities at all levels of ability. We find higher relative marginal earnings 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 earnings. Moreover, we find strong complementarities between knowing computers and studying science, suggesting that computer fluency is important for translating science knowledge into higher earnings. Further, heterogeneity analysis suggests that greater earnings associated with science are concentrated among those who do not have professional degrees.[7] We also find that greater earnings do not accrue, on average, to

[6] Caselli et al. (2014) survey important papers in the literature of Mincerian returns for each country, listing Kingdon (1998) and Agrawal (2012) for India, neither of which fully address endogeneity in years of schooling.

[7] Professional courses refer to the study of engineering, medicine, management, accounting and law.

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 at the impact of studying science on the income, conditional on one of the employment types, we find that while studying increases returns in private employment, among those who are in business, returns from studying science accrue only to those who have high scholastic scores.

As pointed out above, given the lack of 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 interval bounds do not include the null, and that the estimated earnings associated with studying science are robust to potentially large selection on the basis of unobservables.

While the robustness procedure described above may alleviate concerns about omitted factors overturning the main result, we explore the role of additional behavioral factors that might influence students' decision to take science and bias our results.[8] Using a primary survey of 524 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 commonly measured cognitive and non-cognitive skills differ across students who choose science and non-science majors. We find a weak correlation between grit and the likelihood of studying science major. In addition, we find that ambiguity aversion, positive personality traits and Cognitive Reflection Test (CRT) score do not differ by stream choice. Thus, there is no systematic correlation between behavioral traits and the decision to study science.

This paper contributes to a number of strands of the literature. First, the literature

[8] A line of research, summarized by Heckman and Kautz (2012), shows the importance of cognitive and non-cognitive skills factors in the labor market.

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.[9] 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 earnings associated with major choices in a developing country context. The literature on the returns to education largely focuses on the returns to years of schooling , while ignoring the role of majors, language skills, computer skills and other dimensions of schooling. Notable exceptions that unpack the impact of education content include Munshi and Rosenzweig (2006) and Chakraborty and Kapur (2009) on English-medium instruction, Jain (2017) on mother-tongue instruction, and Cantoni et al. (2017) and Dhar et al. (2018) on curriculum.

Finally, our paper relates to the literature on determinants of the stream and occu-

[9] 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).

pational 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 that post-secondary human capital investment is an important determinant of 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 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. At this stage, students must choose from three standard majors or specialization in the academic streams: Science, Business and Humanities.[10] At the end of schooling, students pursue higher education, or enter the labor force. The duration of higher education varies from three to five years depending on the course. Bachelor in Arts and Science programs are three years duration, technical courses are four years, and medicine and architecture last five years.

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*

[10]The state or all-India boards of secondary education determine curriculum at the higher secondary level. The curriculum that a particular school follows is determined by the state or national board to which it is affiliated. For instance, schools affiliated with the New Delhi-based Central Board of Secondary Education will follow its curriculum and offer the Board's All India Senior School Certificate Examinations.

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.

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.[11] However, in practice, the eligibility thresholds are higher for science stream as compared to business or humanities.

There are multiple reasons why students take science in high school in India, apart from heterogenous tastes. One factor is possibly that the science stream and associated career paths are more prestigious. Another possible factor is that an undergraduate degree in science or a professional undergraduate course following science is sufficient for a high quality job, whereas non-science undergraduate programs require further education for similar positions. Moreover, science students can pursue various non-STEM undergraduate courses and careers but not vice-versa, so the number and types of jobs available after studying science in high school is much wider than otherwise.

On the costs side, students studying science might supplement school education with private coaching to prepare for entrance examinations to elite engineering and medical colleges.[12] Finally, students might migrate for better schools or private coaching, with associated financial and non-financial costs.

3 Data

The main empirical analysis uses the India Human Development Survey (IHDS) data collected in 2011-12. IHDS is a nationally representative, multi-topic household survey cov-

[11] 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.

[12] Azam (2016) reports that the average cost of 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.

ering 42,152 households across India.[13] 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 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 in higher secondary stage (grades 11 and 12), and report information on both major choices and earnings. We do not include women in the analysis because women in India have a low labor force participation rate (Klasen, S. and Pieters, J., 2015; Pande and Moore, 2015). 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 whom report being in wage or salaried employment, in business, or being self-employed. Only 4% report being unemployed. Since they earn zero income, we drop them as we use log of earnings as our primary dependent variable.[14]

The IHDS reports household level enterprise profit along with the labor time contribution in the enterprise of household members. We use this information to calculate earnings for business/self-employment.[15] 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 labor is typically inelastically supplied by most male adult members of households in developing countries like India, the actual amount of work is likely to reflect demand for

[13]The survey covered all the contiguous states and union territories of India. For data analysis, we use IHDS design weights to obtain nationally representative statistics.

[14]Our results are robust to including the unemployed using a dependent variable in levels.

[15]Typically, 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.

labor. Thus, the demand for labor 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 labor market, the fact that most public employees are assured work throughout the year, ensures larger earnings from such jobs.

Table 1 reports descriptive statistics for a number of relevant variables. The first row reports the mean annual earnings are Rs. 178,330 (approximately, $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, 36% of the sample speaks fluent English, 48% speak English less fluently while the remaining 16% 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.[16]

Given the background of those surveyed, is there a difference in earnings between those studying science and those studying other majors? Figure A.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 (Figure A.2) as compared to those with lower divisions (Figure A.3).

[16]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.

Finally, in order to understand other correlates of studying science major, we conducted a primary survey on grade 12 students in six districts across Bihar and Andhra Pradesh states in India. We chose these two states since they have a high proportion of science students while varying on a range of other economic and social dimensions.[17]

The survey was conducted at the beginning of the academic calendar (May to July) in 2017 in district towns of Patna, Bhagalpur and Sitamarhi in Bihar, and Vijayawada, Kurnool and Srikakulam in Andhra Pradesh. Schools in these towns were randomly chosen, stratified on private versus public management. Students across science, business and humanities majors were chosen at random from school lists and interviewed at home. We also interviewed one of the parents, so we only surveyed students who co-resided with at least one parent.[18]

While we do not make any claims of representability for the nation or the state, the sample is representative of high school students in these specific cities in India. The population of Patna in Bihar and Vijayawada in Andhra Pradesh exceeds one million while the remaining four are mid-sized cities. Apart from usual idiosyncratic differences, to the best of our knowledge, these cities are not demographically, culturally or economically distinctive.

The final sample consists of 524 students (matched to their parents) in class 12 (the last year of high-school) across 55 schools spread across the two states. Apart from information on subjects chosen by students and their life and career aspirations, most importantly, the survey also measured various behavioral parameters: grit, ambiguity aversion, cognition-reflection ability and positive personality traits, typically not available from developing countries (see detailed descriptions of these in Appendix Table A.2). Appendix Table A.3 presents summary statistics for the survey sample.

[17] Of students of age 16-18 years and who attend school, 65% study science in Bihar and 91% in Andhra Pradesh, in contrast to 55% for the entire country.

[18] Students in India often reside with extended family members, 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; it is less likely for extended family to influence major choice or career decisions of students.

4 Empirical Analysis

4.1 Specification

We use the 2011-12 round of the IHDS with one observation per individual, and estimate the earnings associated with studying science in high school using the following specification.

$$y_{id} = \beta_0 + \beta_1 Science_{id} + \beta_2 \mathbf{X}_{id} + \lambda_d + \epsilon_{id} \tag{1}$$

In equation (1), the main outcome of interest y_{id} is log earnings of an individual i residing in district d. The variable $Science_{id}$ is 1 if the individual studied science in high school, and 0 otherwise. The primary coefficient of interest is β_1, which is the percent increase in earnings associated with studying science in high school. We add a vector of control variables, represented by $\mathbf{X}_{id}$, 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.[19] 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.[20] IHDS data permits controls for ability using individual performance on the secondary school certificate (SSC) examination conducted at the end of grade 10. In order to control for ability among the less educated, we add whether the individual ever failed or repeated a grade. The specification controls for average household education (excluding the respondent) to proxy for household level ability, as well as for parental education, a traditional control to proxy for ability in the returns to education literature (Card, 1999).

Fluency in English might directly effect labor market returns (Azam et al., 2013), so

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

[20] According to the authors calculation using the estimation sample, 39.31% of students who receive first division (higher ability), 20.48% of students who receive second division, and only 10.22% of students who receive the third division choose to study science in grade 10.

$\mathbf{X}_i$ includes measures for self-reported English fluency, represented by indicator variables for "very fluent", "little fluent" and "not fluent". We add a 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_{id} represents $i.i.d.$ unobserved factors that might influence earnings, and is clustered at the state level.

In addition to log earnings, we estimate equation (1) for a number of follow-on outcome variables. These variables represent human capital achievement (specifically, years of schooling, whether the individual completed graduate education, and whether the individual completed professional education) and employment (in particular, public sector tenured employment, private sector tenured employment, and employment in business, as well as income associated with public, private and business employment).[21]

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 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 3% greater earnings.[22]

Examining heterogeneity in the results helps to determine the pathways through which

[21] Private sector tenured employment represents employment within the private sector with self reported security of tenure. When we consider income associated with private employment, we include tenured as well as untenured private sector employees as the income gains may be correlated with getting tenure.

[22] Appendix Table A.1 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.

science education transmits to earnings. 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 the top 1%. At the 99^{th} percentile, we find that studying science in high school is associated with 37% greater earnings. Thus, our results illustrate an important driver of convexity in returns to education by highlighting that stream choice is correlated with the highest incomes.

Higher ability students might be, 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 earnings associated with science is 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 workers.

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 earnings associated with different skills. If STEM jobs require extensive communication with others, especially in the business world where language skills are important, then the labor market value of science education might be influenced by English fluency. Conversely, if STEM careers require expertise in science with communications handled by other employees, then earning of workers with science proficiency would be independent of their language skills. For similar reasons, the value of science 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 earnings from science accrue significantly more when an individual knows English. The earnings are 28% greater with fluency in spoken English ($p < 0.01$), and 19% higher with little English ($p < 0.01$). The earnings associated with 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 science education in the job market.[23] Mirroring these results are the findings associated with computer skills in Panel B. Science education is associated with high earnings (+31% for first division students, $p < 0.01$; +19% for second and third division students, $p < 0.05$), especially when the respondent was proficient in computers. Earnings are significantly lower (7%, $p < 0.10$) for students who report they are not proficient in computers. Collectively, these findings point to the critical role of communication and technical skills in operationalizing greater earnings from science education.

The returns to education literature for India shows that market oriented professional courses in engineering, medicine, business, law and accounting yield the greatest earnings (Duraiswamy, 2002). Such courses typically command higher wages as compared to "general" university education, partly because their skills sets can be readily deployed without firm-specific training. Therefore, complementarities between science education and market oriented skills are interesting to investigate. Panel C of Table 5 finds that science education has no impact when comparisons are made among those who have completed professional degrees across all ability levels. 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. In contrast, we find studying science

[23] Our results on complementarity is consistent 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 two-thirds to three-fourths of the differential in earnings growth between immigrant and native employed in high-skilled occupation.

have higher earnings across all ability levels among those without such market skills, i.e., those who did not complete a professional degree. This finding suggests market oriented degrees and science education in higher secondary school are potentially substitutes in the labor market.

We next examine how the social environment, represented by social group and parent education, influences the value of studying science. Socially privileged individuals might benefit disproportionately more from science 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 earnings are 25% greater 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%, $p < 0.01$).[24] This pattern holds when examining by ability subsamples specified earlier. Combined, the results in Table 5 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

[24]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).

earnings.

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 can be classified into one of four types of employment: a position in the private sector with security of tenure (we refer to it as Private Tenured Employment), a relatively secure job in the public sector, running one's own business and untenured jobs largely in the private sector. 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 tenured 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, private jobs are competitive at the top end of the wage distribution. 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 obtaining a job in that sector.

Table 8 shows the implications of science education on income associated with employment in different sectors, conditional of being selected in a particular kind of job. We include both tenured and untenured jobs in the private sector, because security of tenure and income are likely to highly correlated in the private sector. This is not so important for the public sector as most (94%) public sector jobs are already tenured (permanent). Panels A, B and C report that earnings from science are the highest among high ability students who operated their own business (+0.42, $p < 0.10$). Earnings from science versus non-science are for high ability students not very different in the private (+0.21, $p < 0.10$) and public sectors (+0.26, $p < 0.01$). For those with lower scholastic ability, science offers little earnings boost in business ownership and 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.

4.4 Robustness

Though the main regression model controls for ability by including dummy variable for divisions, a concern is the possibility for other kinds of unobserved abilities not completely subsumed by scholastic performance, as well as households factors to potentially bias our estimated results. In this section, we assess the extent of potential bias due to the exclusion of these variables in the model following the 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_s because of W. Altonji et al. (2005) estimate this bias by assuming the following.

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

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 keeps account of movement in the R-squared value due to addition of control variables. Following this method, we 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, we start with a baseline regression where log of earnings is regressed on $Science$, and then add further controls. As a second step, we 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 the literature contains Mincerian returns to education regressions for India, none look at the earnings of urban males who have completed high school.[25] Given the lack of a known R_{max}, we follow Oster (2017)'s suggestion and set R_{max} as 1.3 times the R-square of the regression that controls for Z (controlled regression). Since the R-square in our main specification is 0.304, we set $R_{max} = 0.4$. 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 $\beta_s(0.4, 1)] = 0.16$. Moreover, we provide the value of δ for which β_s would become 0. The obtained value of 3 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, this exercise indicates that the estimated earnings associated with science education are robust to potential omitted variable bias.

5 Role of behavioral characteristics

Recent literature has highlighted the role of behavioral characteristics in the labour market. While Section 4 shows that the biases due to the omission of these characteristics may not be large, we delve explicitly into how strong is the correlation between behavioral characteristics and the choice to study science. Drawing on a unique data from a survey of high school students in the states of Andhra Pradesh and Bihar (described in Section 3), this section explores if there is any role of some of these factors in students' choice of science.

Our primary interest is unpacking the role of behavioral characteristics in students' decision to study science. These characteristics are generally unobserved in most studies in the developing countries. Our survey includes measures of students' grit, ambiguity aversion, cognitive ability, and personality and non cognitive skills, typically not available in any nationally representative data sets (see descriptions of each in Appendix A.2). These

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

are supplemented with information on individual and household characteristics, from both student and parent respondents.

The literature has pointed to individual grit being correlated with educational success and other long-term goals (Bowman et al., 2015; Duckworth et al., 2007), even after controlling for IQ and Big Five conscientiousness. Thus, grit could be important in the labor market and subsequent economic returns. We investigate the role of grit in the decision to study science by using standardized questions suggested by the psychology literature, and scoring each sampled students on the grit scale (Duckworth et al., 2007).

In order to make fully-informed education choices, students should have information on earnings associated with different types of jobs and also know the probability of obtaining such jobs. However, our qualitative survey revealed that students and parents had poor knowledge at the high school stage of options that follow from studying different subjects.[26] One possibility is that students may not want to make substantive choices till they have better information on labor market options. Since studying science leaves options open to study all subjects whereas studying business or humanities forecloses STEM options, the choice of high school major might be correlated to students' willingness to make decisions in ambiguous situations. Hence, those who choose to study science might be relatively more ambiguity averse. To investigate this hypothesis, we measure ambiguity aversion using ambiguity tolerance scale suggested by the psychology literature (MSTAT-II) as well as ambiguity experiments suggested by Ellsberg.

The expected labor market value from a science education is a potentially important determinant of subject choice. Greater cognitive ability facilitates understanding the relative costs and benefits associated with this decision. We use a three-item Cognitive Reflection Test (CRT), suggested by Frederick (2005), as a measure of cognitive ability. This measure is predictive of the types of choices that are used to test expected utility theory and prospect theory. Finally, the recent literature (Heckman and Kautz, 2012; Acosta et al., 2015; Deming,

[26]For instance, many students and parents were unable to name their dream institutions after high school nor the kinds of jobs that could follow.

2017) emphasizes the role of socio-emotional skills (personality traits and behaviors) in the labor market. Consistent with this, we collect information on students' personality traits.

Table A.3 describes several variables from the survey dataset. We collect information on major choice in high school, and 60% of students were enrolled in the science stream. A large fraction of students (62%) earned first division in grade 10, reflecting perhaps the concentration of more capable students in urban schools, or the selection of more able students into the higher secondary stage. Survey responses indicate that students reflected on their stream choices (75%) and that challenging careers and subsequent earnings are important for them (80% and 83%, respectively). The responses also indicate that 43% of parents' are involved in their children's educational choices. Siblings are more frequently sources of career information compared to friends (41% versus 28%).

Table 10 summarizes the effects of the behavioral measures described above on students' decisions to take science. The full specification in Column (2), which includes the full set of control variables, shows that the grit score is positively correlated with the choice of science (+0.10, $p < 0.10$). Greater ambiguity is also negatively correlated with taking science, although the ambiguity score coefficient is relatively small (-0.006, $p < 0.10$), and the ambiguity experiment score is not statistically significant (-0.085, $p > 0.10$). Neither the CRT score nor the personality variable have significant correlation with science education. Overall, we interpret these findings as suggestive that there is, at best, weak correlation between these commonly measured behavioral characteristics and influences on educational choices at the senior secondary stage in the context of our sample.

6 Conclusion

We explore the role of science education in high school, a stage where important career choices are made, on subsequent education, career and earnings outcomes. Our analysis, though not causal, shows that science education at the higher secondary level 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.

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. Causal estimates of the effects of science 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. Second, due to data limitations, we include neither women nor rural residents in our sample. Since the dynamics of how science education translates into earnings and employment is potentially very different for these groups, we caution against extending our estimates for these groups. Finally, we do not analyze potential barriers to the effectiveness of different pedagogical approaches to science education, which might create significant variation in the estimates associated with professional outcomes. We hope that these issues will be addressed in future research.

Nonetheless, our analysis informs the global debate over the value of STEM-focused education versus a traditional liberal arts curriculum. Our findings not only suggest high labor market earnings associated with high school science, but also suggest strong complementarities between science and other skills such as computer and English fluency. Collectively, results from this paper point to the importance of policies to facilitate the ability of socially disadvantaged individuals to undertake human capital investments. ■

References

Acosta, P., N. Muller, and M. Sarzosa (2015). Beyond qualifications: Returns to cognitive and socio-emotional skills in Colombia. World Bank Policy Research Working Paper No. 7430.

Agrawal, T. (2012). Returns to education in India: Some recent evidence. *Journal of Quantitative Economics 10*(2).

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. University of Texas working paper.

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.

Cantoni, D., Y. Chen, D. Yang, N. Yuchtman, and Y. Zhang (2017). Curriculum and ideology. *Journal of Political Economy 125*(2), 338–392.

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.

Caselli, F., J. Ponticelli, and F. Rossi (2014). A new dataset on Mincerian returns.

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

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 132*(4), 1593–1640.

Dhar, D., T. Jain, and S. Jayachandran (2018). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India.

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. (2017). Common tongue: The impact of language on educational outcomes. *Journal of Economic History 77*(2), 473 – 510.

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.

Kingdon, G. (1998). Does the labour market explain lower female schooling in India? *Journal of Development Studies 35*(1), 39–65.

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? *World Bank Economic Review 29*(3), 449–478.

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*, 1–18.

Pande, R. and C. Moore (2015). Why aren't India's women working? *The New York Times, August 23*.

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:		
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 Tenured Employment	0.25	0.43
Dummy: Public Tenured 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 four categories of employment are Business Employment, Public Tenured Employment, Private Tenured Employment with Non Tenured-Non Business Employment as the omitted reference group. The number of observations for the variables Average Household Education and Max Parent Education are 4,687 and 2,513 respectively.

Table 2: Earnings and 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: Earnings and high school science major, by quintile

	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: Earnings and 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 earnings associated with 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: Science majors and 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: Science majors and employment outcomes

	Full Sample	Division I	Division II/III
	(1)	(2)	(3)
Dependent Variable:	PANEL A: Public Tenured 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 Tenured 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: Science major and income

	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	0.24***	0.21**	0.29***
	(0.05)	(0.08)	(0.07)
Constant	3.26***	4.03***	3.32***
	(0.40)	(0.84)	(0.48)
R-squared	0.38	0.44	0.40
Observations	2,143	661	1,482
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 employed individuals (both tenured and non tenured), all private employed individuals (both tenured and non tenured) 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 $\beta_s = 0$	R^2_{max} for $\beta_s = 0$
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: Behavioral correlates of high school science major

Dependent variable: High school science major		
	(1)	(2)
Grit score	0.077	0.010*
	(0.051)	(0.054)
Ambiguity score	-0.003	-0.006*
	(0.003)	(0.003)
Ambiguity experiment score	-0.082	-0.085
	(0.078)	(0.074)
CRT score	0.067**	0.032
	(0.028)	(0.022)
Personality	-0.006	-0.001
	(0.009)	(0.009)
Household controls	No	Yes
Demographic controls	No	Yes
Other controls	No	Yes

NOTES: Data from primary survey conducted by authors in 2017. Dependent variable is 1 if student chose science in class 12. Household controls in all regressions include household size, mother completing class 10, father completed class 10, asset index, distance to closest bank, father salaried employee, mother salaried employee. Demographic controls include age, caste, religion (included gender in the full regression but not in these ones which have male only sample). Other controls include city tier and state board syllabus. Errors clustered at school level.

Figure A.1: Earnings distribution by high school major

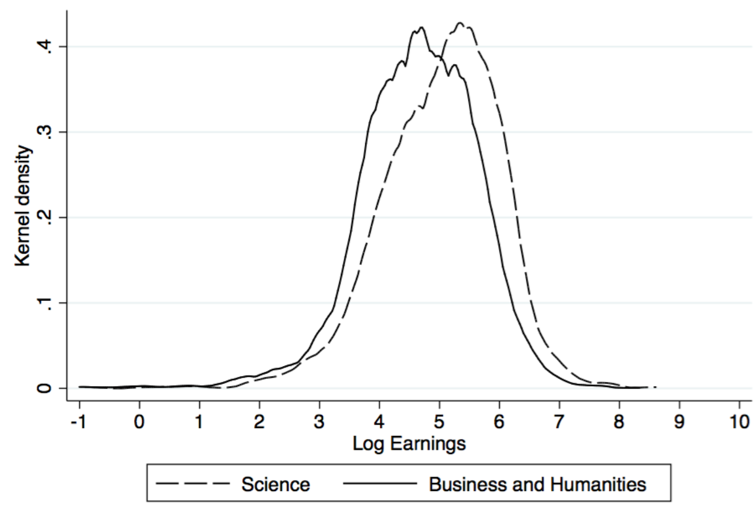

Figure A.2: Earnings distribution by high school major - First division students

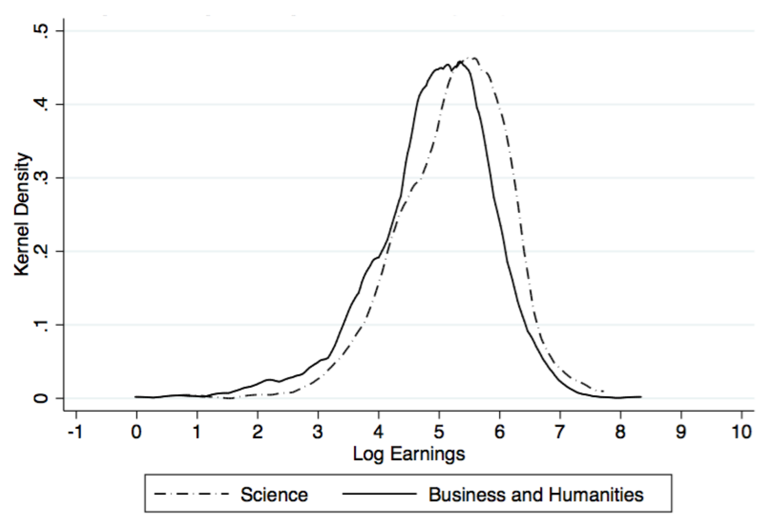

Table A.1: Science majors and earnings (including unemployed individuals)

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. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure A.3: Earnings distribution by high school major - Second and third division students

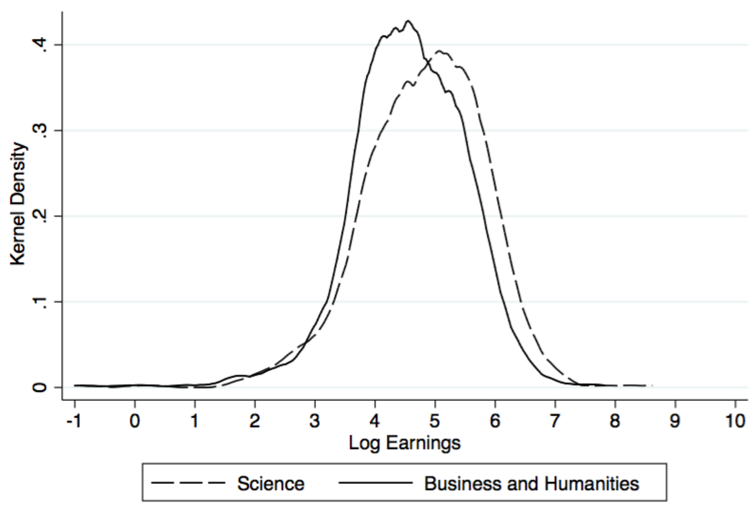

Table A.2: Description of variables

Name	**Description**
Grit	*Grit* is defined as the perseverance and passion for long term goals. We employ Duckworth et al. (2007)'s 12-item Grit Scale. During the survey, respondents rated 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'. A high score on the aggregated Grit scale indicates 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 respondents was measured using the Multiple Stimulus Types Ambiguity Tolerance Scale - II (MSTAT II). This 13-item psychometric scale assesses the cognitive response of respondents to different ambiguous stimuli McLain (2009). 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	Respondents 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. Respondents then picked 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 is an adaptation of the famous "Ellsberg Paradox" in which participants were found to prefer situations with known probabilities of events to unknown probabilities of events.

CRT score	*Cognitive Reflection Test (CRT)* is a test of how quickly respondents process and respond to basic aptitude questions, ignoring an obvious looking incorrect answer, and instead processing the question and responding with a correct answer. Each correct answer was awarded one point, with a total score for each student calculated out of 3. The questions are as follows. • 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?
Personality	This test measured respondents' personality and non cognitive skills. Respondents rated their agreeableness on a 5 point scale (with 1 corresponding to 'Do not agree' and 5 corresponding to 'Completely agree') to a set of positive statements related to personality and non cognitive skills. These scores were aggregated for all statements to create a cumulative personality score. The questions are as follows. • 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.

Table A.3: Survey data summary statistics

Variable	Obs.	Mean	Std. dev.
Academic measures			
Science	524	0.60	0.49
Math score in class 10	373	70.10	19.02
Science score in class 10	366	68.50	16.93
English score in class 10	301	70.83	20.18
First division in class 10	524	0.62	0.49
Second division in class 10	524	0.14	0.35
Third division in class 10	524	0.13	0.34
CBSE syllabus in class 10	524	0.23	0.42
ICSE syllabus in class 10	524	0.11	0.31
State syllabus in class 10	524	0.65	0.48
Behavioral characteristics			
Grit score	524	3.43	0.62
Ambiguity tolerance score	524	40.29	8.85
CRT score	319	0.76	0.96
Personality score	524	31.34	3.35
Other variables			
Student gave a lot of thought on his/her stream choice	524	0.75	0.43
Student thinks science stream is for smarter students	524	0.27	0.44
Challenging career is important for student	524	0.80	0.40
Earnings is important for student	524	0.83	0.37
Career with travel opportunities is important for student	524	0.65	0.48
Career that allows to stay in big city is important for student	524	0.70	0.46
Career that emphasizes managerial skills is important for student	524	0.56	0.50
Career that has non-transferable job is important for student	524	0.52	0.50
Parent gave a lot of thought on student's education	524	0.43	0.50
Parent thinks stream choice is important signal	524	0.59	0.49
Parent thinks stream choice is important for job	524	0.52	0.50
Friend took science	524	0.73	0.45
Friend took commerce	524	0.20	0.40
Friends took arts	524	0.11	0.31
Referred to siblings for information	524	0.41	0.49
Referred to friends for information	524	0.28	0.45
Household characteristics			
Student age	520	16.93	0.96
Bihar	524	0.61	0.49
Mother completed class 10	524	0.67	0.47
Father completed class 10	524	0.85	0.36
Household size	524	4.84	2.00
Distance to closest bank (in kms.)	523	2.41	5.26
Religion: Hindu	524	0.89	0.31
Religion: Muslim	524	0.06	0.25
Scheduled Caste	524	0.16	0.37

General Caste	524	0.32	0.47
Other Backward Caste	524	0.51	0.50
Tier I city	524	0.41	0.49
Tier II city	524	0.35	0.48
Tier III city	524	0.23	0.42
Electric connection	524	0.98	0.14
Land line telephone	524	0.04	0.21
Internet connection	524	0.35	0.48
Tap water supply	524	0.69	0.46
Student has access to cell phone	524	0.68	0.47
Student's phone has internet access	514	0.45	0.50

NOTES: Data from primary survey conducted by authors in 2017.