Education Attainment of Households Based on Scholarships in India: Evidence from 2014-2018

Beard Seth, Broyles Thatcher, Dongre Rohit, Walthare Rohit

December 9, 2022

1 Introduction

Around the world, educational policies and programs are created and implemented to make it possible for underprivileged students to access great education. Though the percentage of illiteracy has fallen in recent decades in some parts of the world, it remains in the domain of unexplained. It is unexplained because despite witnessing good provisions, laws, and policies over the years, it has had a superficial impact on the communities those policies had been enacted for. In this context, India is an interesting case study. It is interesting because here the Caste system significantly affects every core of life. No matter its social, political, educational, or economic despite having inhuman practices based on Manu-smriti (doctrine of laws) Caste System of India has fewer critiques and more followers. Every day it affects individuals in their public and private spaces. In most cases however, these impacts are unobservable which creates the problem of selection bias. For the same, academic education is no exception to it. We are controlling the caste factor in our study to see the education outcome in public and private institutes based on the scholarship program in India.

Some recent studies indicate that public schools are deficient in quality education. The reasons to advocate these studies are often bad test scores, poor infrastructure, underqualified teachers, and teacher absenteeism. However, the contradiction is if public institutes are really deficient in providing quality education, then it must be true in the case of college or

university-level education offered by the public institutes. But that is not the case. Because when it comes to university-level education millions of students every year compete for getting into public universities which are elite and prominent institutes of India that are altogether sponsored by the government.

If a student from a lower caste and a student from a higher caste gets the same entry score, the student from the higher caste will get before the lower caste student. This caste problem finds its way into both public and private education. If people or governments find flaws and disparities in the functioning of public schools, then they can put better regulations and control mechanisms intending to make it efficacious. Nevertheless, there might be other unobserved characteristics that may undermine the performance of the public school, or it could be because of poor governance. However, in every case, the public schools must be fostered and disseminated and not the other way around. In the era of the commercialization of education, we express interest in the public schooling reform of education in India.

1.1 Importance of public education

The Indian subcontinent is an amalgamation of a diversity of populations segregated based on different castes and religions. Considering the economic and demographic characteristics of India here model of commercialization of school education cannot flourish. To understand this, we have to take caste into account. In India, the socio-economic characteristics of the population are contingent and vary across caste groups. It is like peeling the onion and unfolding the problems associated with it. It also coincides with the problem of race, however, the story of caste lies beneath the skin, and, so, the caste problem is vast and capable of much wider mischief both theoretically and practically. Because poverty and marginalization in every field decide by the stratification of social groups.

If individuals belong to the superior caste, it is highly likely that they would get more opportunities for good education. Let's say private schools provide a better quality of ed-

ucation, even though individuals from the Dalit caste are equally competent as somebody from a high caste, they will not be able to enter into the private school probably because they may not have enough resources and money to afford the cost of private schools. This case relates to omission sensitivity which economists often discard. If ones are equally likely to enter into private schools, and their scarcity of resources prevents them, we cannot merely advocate that the test scores are enough to decide the quality of education in public schools. It is not the test scores it is also about the circumstances that they live in. Practically there is no universal way to calibrate an individual's intellectual potential. This is mainly because the value of his and my test scores cannot be the same even though we have given the same exam at the same age during the same time period, and I scored more than him. I have resources to study while my friend doesn't. How can his value of score be the same as mine? He is intelligent even though he scores a B and my A would be inferior to his B. My score cannot be equivalent to his efforts. Because I have privileges of my priestly class. Hence, they can experience equal opportunities only when public schools are fostered.

1.2 Importance of Scholarship

Scholarships are a way to empower students from marginalized communities. When we talk about India and its social group scholarships often work as catalysts for these sections of society. In the earlier section we described the circumstances, they live in such problems can also be addressed by providing scholarships. Because scholarships in India are not merely given on the basis of merit but also on the basis of historical exclusion and persecution of these groups. The students who get scholarships go beyond and excel better in life. The majority of these populations attend public schools because they cannot afford the high standards of private schooling.

2 Literature Review

When looking at education attainment in Indian society, there were a couple studies previously done that relate to the proposed question in our paper. The first study "Does Affirmative Action Work? Caste, Gender, College Quality and Academic Success in India"(Bagde, Surendrakumar) discussed public policy and the outcomes of affirmative actions programs in India regarding quotas instituted by the Indian government. This paper shares a similar theme with our question looking at scholarship and affirmative action. The question of interest looks at a specific quota program and the effects it has on college admission and attendance for students of disadvantaged castes and women. The program outcomes showed an increase in college attendance of women and lower caste students especially at top tier institutions. The study first points out the Quota System that is currently in place due to the law of the Supreme Court of India. The following percentages are the allotments of seats(spots) from disadvantaged social groups in top universities:

- 15 percent for Scheduled Castes
- 6 percent for Scheduled Tribes
- 7 percent each for BC-A and BC-D
- 10 percent for BC-B and 1 percent of BC-C [1]

The study explains the process for seat occupancy in university programs going off of entry test scores. Priority for admission is based on students rank on a country wide entrance exam. If the individual ranks the highest in their caste, then they are able to attend any school they want in order to fill the quotas. Once the quotas are full, spots move on to men starting in higher castes and then make their way down. Women are also offered spots but after men of the higher castes pick their schools. The research uses an empirical model which shows the probability of engineering college matriculation using the following variables: student, caste, and effective rank. Effective rank is a variable plotted against "Entry Exam Rank" to show how the students would rank taking caste and gender into account.

[1] BC is abbreviated for Backwards Caste and subdivisions in it.

This results in the Open group with the lowest effective rank and the Scheduled Tribes having the highest effective rank. This linear probability model then produces the an increase in the output variable of college admission and attendance of these lower caste students. In our research, we would ideally want to have test score data to further our regression models but unfortunately it was not available. Our research design can be seen a deviation from this model in the sense that we are using similar data but are looking specifically at educational attainment over time.

In the Journal of Development Economics, the study "Affirmative action, education and gender: Evidence from India" (Cassan, Guilhem) question of interest looks at the "impact of India's affirmative action policies Scheduled Casts on educational Attainment" (Cassan, Guilhem). The study finds that affirmative action does in fact increase educational attainment with large gains in primary education, secondary education and literacy among men but not women. Main assumptions in this study include that there is no migration across regions where data was collected and that there is no identify manipulation within the units of the data set. However, relating back to the first article mentioned, these benefits were only seen for the male gender.

The paper illustrates the differences between Scheduled Castes, Scheduled Tribes and Backwards Castes just as in our model where we categorize each caste. A difference in our study is that in this existing research paper, they include age in the data and only include units above 18 years of age. Meanwhile in our analysis, we include age as a descriptive statistic that ranges from 3-35 years of age. This statistic ends up being statistically significant in our study based on p-value. The researchers ran double and triple difference estimations in order to separate effects of the affirmative action between castes and gender. Increases in literacy, primary and secondary education were seen in the double difference estimation and even slightly higher increases in primary and secondary education in the triple difference model. There was no evidence of an increase in higher education however among men.

Women did not see an increase in any of these variables and instead even saw some of these statistics decrease after the double and triple difference estimations. The paper ends up focusing more on the difference in effects between men and women showing that women in some cases did the opposite of benefit as some of the coefficients became negative during the regression. Our model will focus more on the other side of this paper which is the caste system and how certain states/districts with higher literacy and education rates compare to those states/districts with lower literacy and education rates.

3 Data

We used National Sample Survey (NSS) data from the National Data Archive of India, which is kept up to date by the Ministry of Statistics & Programme Implementation. Data surveyed by the Government of India from the years 2014 and 2018. The data provides the demographics of the households as well as the total number of residents. It is also the first time the Indian government has conducted a survey based on the household's educational attainment. This survey includes all of the Indian states as well as the Union territories, albeit certain villages in the Union territories are left out because of the geographical location of islands and access issues. However, sample selection and allocation in surveys were done by listing the total number of the first stage units of households to the States and Union Territories in proportion to population as per Census 2011 subject to a minimum sample allocation to each State/UTs.

The survey was conducted in four stages, including listing the households, household consumer spending, household social consumption, health, and education. This survey is the primary source of statistical indicators on social consumption and well-being, the standard of living, and related inequality, as well as for estimating the number of other parameters. However, the vital component of the survey is the relevance to the determination of the prevalence rate of morbidity among various age-sex groups in different social and other

groups in different regions of the country along with the household social consumption on education. This survey covers both the qualitative and quantitative aspects related to the education and health of the household members.

We used the data reported by this survey on both rounds, i.e., 71st and 75th. Given that these social groups make up the majority of the people in India, it was intriguing to see how the general level of education changed as a result of the necessity to provide scholarships based on the social status and educational disadvantages among these groups in the past. However, while offering scholarships to various social groups, we wanted to know how the level of test scores changes with educational attainment. Nevertheless, due to data limitations, this analysis was left out of this study.

The variables such as level of income, monthly expenses, the distance to the school (primary/secondary or high school), and other factors that could have caused endogeneity in the data were omitted in the analysis. Scholarship, socioeconomic groups, institution type, religious status, free education received, textbooks/stationery received, and a midday meal was the primary variables of interest for the analysis. The social group was the main parameter because scholarships were typically received by this group based on social/economic status rather than solely on merit. This group includes the four categories of Schedule Tribe, Schedule Caste, Other Backward Class, and Others also, the survey used two groups to determine whether or not a person received a scholarship.

Additionally, the data includes the statedistrict variable, which depicts households at the district level from each unique state and we found 681 distinct clusters at the district level for the analysis. When we merged the two datasets for both survey years, we had more than 0.8 million observations and 200+ variables, but for the analysis of this study, we only ended up using 21 variables and 246,505 observations that were relevant to the study's objectives.

4 Descriptive Analysis

The variables between the 71st and 75th rounds are identical when examining the descriptive statistics. The average general education level of people enrolled in school for one year to sixteen years is the first statistic we compute. We found that the average in Round 75 was marginally higher than in Round 71, i.e., roughly 8.35 years to 8.23 years, respectively. Between the two data rounds, the age of the study's units, i.e., people, ranged from 3 to 35, with both means being calculated to be just over 13.8 years old. The gender of each member of the household is the next variable we take into consideration. We decided to strictly focus on the impact of scholarships and social groups on educational attainment and not on gender inequality, even though the existing literature we discussed earlier in the paper runs a lot of regression models on the difference in educational opportunities between men and women.

The code for male was one, whereas for female it was two, and any other gender was classed under three , which is only covered by round 75th. Despite this, we kept it as a variable in our regression models. The gender mean was under 1.50 in both years, indicating that there are more men in the sample than women. When examing the scholarship variable, it indicates whether a student received a scholarship on an individual basis. The percentage of the overall scholarship obtained by the person on a district level is presented by the variable we created for the study called the scholarship ratio. Let's say that, out of ten people, five received scholarships and five did not.

The output from the scholarship ratio variable would be 50%. When comparing the two sets of data, the 71st round of scholarships results in a 1.81 mean, where one denotes receiving a scholarship and two denotes not receiving one. This indicates that in the 71st round, around 19% of students did get a scholarship of some kind. Similar to the previous data set, the 75th round yields a mean of 1.84 for scholarships given, indicating that around 16% of students in this data set were given scholarships. The 71st-round mean for the

district-level scholarship ratio is determined to be 0.18, and in the 75th round, it is around 0.15. This shows that a higher percentage of the districts gained a scholarship in the 71st round compared to the 75th round.

The following are the last four descriptive statistics: free education, textbooks, stationery, and a midday meal. If the individuals in the regression acquired any of these advantages in terms of change in education level, that is determined by all these statistics. With means exceeding half of the dummy variables, more than half of the students in the regression obtained each of these advantages. The student's social group status is the subject of the descriptive statistics section's final section. As indicated by the data, we are concentrating on four distinct social groups among these groups: *1. Scheduled Tribes 2. Scheduled Castes 3. Other Backward Class 4. Others*

From Table-1 - Table-6 we are looking at the summary statistics along with specifics of scholarship received by these social groups. Out of all four social groups, in both the 71st and 75th rounds, Scheduled Tribes had the highest percentage of scholarships received at 32.2 percent and 27.9 percent respectively. The group with the lowest scholarship percentage was the Others group which received scholarships in the 71st round at a rate of 6 percent and in the 75th round at a rate of 5.4 percent. From the comparison of the figures mentioned, could have seen that disparity between people receiving the scholarship is mostly basis on social status and not only on the merit basis i.e. The scholarship is how disadvantaged social groups are able to attend universities and higher education Table-3 - Table-6 shows the breakdown of these statistics based on the institution type attended by these social groups in the year 2014-2018.

5 Research Proposal

5.1 Methodology

The data sets from both rounds of the surveys were cleaned and merged together into one data set using the unique household and personal identification numbers found in each. In order to successfully merge demographic and educational data within each survey, we generated a new variable that combined the household identification number and the personal identification number for each observation into one string variable.

We generated a new variable, scholarship ratio, using the proportion of individuals within each district who received a scholarship. This percentage value was assigned to each district member regardless of whether or not they received a scholarship. The reason we created this variable is to work around the data obstacle that household ID is not consistent across the two surveys, meaning different households may have been assigned the same HHID across surveys. By measuring the proportion of individuals who receive scholarships within a particular district, we are able to isolate the effect of increasing this proportion on years of education.

We determined to use two regression methods in our design, OLS and difference-in-difference. Equation 1 outlines our model for the OLS. We include our primary dependent variables, Scholarship Ratio and Social Group, along with our control variables (age, sex, etc.) In order to control for the endogeneity within the scholarship ratio variable, we clustered the analysis on the district level. Our second model is a difference-in-difference model and is defined in Equation 2. We use scholarship ratio as our treatment variable and the survey round (71st or 75th) as our time control. In the same manner as our OLS model, we cluster on the district level to control for endogeneity.

$$Y = \beta_0 + \beta_1 * ScholarshipRatio + \beta_2 * SurveyRound + ... + Covariates + \epsilon \quad (1)$$

$$Y = \beta_0 + \beta_1 * Interaction + \beta_2 * ScholarshipRatio + \beta_3 * SurveyRound + ... + Covariates + \epsilon \quad (2)$$

In our equations, Y represents the individual years of education attainment. Our interaction variable is defined as ScholarshipRatio*SurveyRound.

5.2 Empirical results

Table 7 reports the OLS regression output from our analysis. Our primary dependent variable, scholarship ratio, has a positive coefficient of 0.08pp and is significant at the 10 percent level of significance. This can be interpreted as for every 1 percent increase in the district-level scholarship rate there is an increase of 0.08pp years of education to the district-level average education attainment. Our secondary independent variable, the caste system, is also significant and has the expected negative coefficient. Our control group for the social groups is individuals not in the lower castes (other). Membership in the scheduled tribe or caste resulted in a negative effect of -0.08pp and -0.05pp years of education respectively. Both coefficients were significant at the 1 percent level of significance.

Table 8 reports our results from the difference-in-difference analysis. The interaction term's coefficient was highly significant (at 1 percent level of significance) and had a positive effect of 0.29pp. This coefficient can be interpreted to mean an increase in scholarship rate of 10 percent results in a .029 increase in years of educational attainment. The caste coefficients were identical to those in the OLS regression, negative and significant on the 1 percent level. The scholarship ratio coefficient was negative, indicating that in survey 71, districts with higher scholarship rates had lower enrollment levels. One possible reason for this observation could be that more scholarships were given to districts with lower educational attainment in

order to improve the average attainment level for individuals.

5.3 Further research possibilities

5.31 *Learn about the similar notion (change in general level of education) based on scholarship intervention and student test results for these social groups.*

The interesting fact that one can comment on the effectiveness of the scholarship programs is the observation of similar results based on the scholarship and individual test scores. Based on this, we could have observed the unobserved characteristics such as whether the scholarship really does make students learn more in terms of getting an additional level of education or affecting their career in the future.

5.32 *This research design could be much more impactful if we had consistent panel data across the two surveys (HHID's match across survey rounds).*

The perfect panel data could be a difficult task to obtain in terms of observing the similar results that we got for this study, and the reason behind obtaining that could lose the number of observations as people move around the places for the job or other opportunities, and it would be difficult to observe the demographics and other variables. Additionally, the four-year interval between our research periods makes it possible that additional events, such as the person's transferring schools or losing their scholarship, may have occurred. But that would be worth mentioning and may have affected the results dramatically as what we believe from the results we received based on the data we have.

5.33 *Although most economists have conducted studies proving that private schools outperform public schools, we argue that this may alter depending on the education policy in each state.*

Since each state has a different set of educational policies and literacy rates, it would be intriguing and informative to learn about outcomes of a similar nature at the state level. That would allow us to draw more insightful inferences from our findings. For the next

research study as an add-on, we plan to explore state-level policies with overall educational achievement based on scholarships and test results.

6 Conclusion

Our paper studies the effects of scholarship rate and caste status on general educational attainment in India. The goals were to show the difficulty of gaining education in the caste system and why scholarship and affirmative action are so important for the lower castes. Scholarship and affirmative action give low-caste students the opportunity to purse an education that will help propel them further in life and in society. As we look at results of both our OLS and Difference in Difference regressions, we can conclude that the variables scholarship rate and caste status were significant on the 1 percent level when looking at educational attainment. These results are promising as in future research we would love to

dive deeper into the variables of public and private to possible show that education policy is the main driver of higher education in each state in India rather than if a school is public or private. Another angle we want to pursue moving forward is scholarship intervention with test scores. Since test scores were not included in the data that we used, we were unable to include that variable in our regression. Diversity quotas in Indian education may look inclusive at first glance, but members of the disadvantaged social groups are still struggling to get an educational opportunities that they deserve. We hope future research targets the education inequality in India and pushes for reform and assistance for those who are less privileged.

Figures and Tables

Table 1: Descriptive statistics for 2014

-> round = 71

Variable	Obs	Mean	Std. dev.	Min	Max
gen_edu	93,513	8.237411	2.389795	1	16
age	93,513	13.80044	5.065153	5	29
Gender	93,513	1.440998	.4965092	1	2
scholarship	93,513	1.819276	.384791	1	2
scholar_ra~o	93,513	.1807235	.1475613	0	.8405797
edu_free	93,513	1.613166	.4870277	1	2
textbooks	93,487	4.2429	2.325745	1	6
stationery	93,474	5.638488	1.194928	1	6
midday_meal	92,844	1.665988	.4716464	1	2

Table 2: Scholarship given in 2014 based on social groups

-> round = 71

	Received scholarship/ Stipend		
Social group	Yes	No	Total
scheduled tribe	4,164	8,733	12,897
scheduled caste	4,622	10,349	14,971
other backward class	6,377	30,661	37,038
others	1,737	26,867	28,604
Total	16,900	76,610	93,510

Table 3: Institution type based on Social groups:2014

-> round = 71

Type of Institution	Social group				Total
	scheduled	scheduled	other bac	others	
Government	7,583	8,657	15,626	10,008	41,874
Private aided	1,380	1,360	4,372	3,550	10,662
Private un-aided	1,554	2,526	10,375	7,907	22,362
Not Known	16	28	66	97	207
Total	10,533	12,571	30,439	21,562	75,105

Table 4: Descriptive statistics for 2015

-> round = 75

Variable	Obs	Mean	Std. dev.	Min	Max
gen_edu	152,992	8.385582	2.766466	1	16
age	152,992	13.84881	5.360607	3	35
Gender	152,992	1.426545	.4946955	1	3
scholarship	152,558	1.845705	.3612329	1	2
scholar_ra~o	152,992	.1538577	.1426517	0	.8333333
edu_free	152,558	1.616756	.4861787	1	2
textbooks	152,558	4.300515	2.304621	1	6
stationery	152,558	5.660595	1.155278	1	6
midday_meal	151,073	1.684709	.4646333	1	2

Table 5: Institution type based on Social groups:2015

-> round = 75

	Received scholarship/ Stipend		
Social group	Yes	No	Total
scheduled tribe	6,083	15,688	21,771
scheduled caste	5,761	19,279	25,040
other backward class	9,316	52,446	61,762
others	2,379	41,606	43,985
Total	23,539	129,019	152,558

Table 6: Scholarship given in 2014 based on social groups

-> round = 75

Type of Institution	Social group scheduled	scheduled	other bac	others	Total
Government	13,880	13,638	26,456	14,716	68,690
Private aided	1,809	1,950	6,575	5,197	15,531
Private un-aided	2,688	4,115	14,622	10,805	32,230
Not Known	47	26	74	103	250
Total	18,424	19,729	47,727	30,821	116,701

Table 7: OLS Regression Output

	(1) Years of education attained
Scholarship_ratio	0.08*
	(0.043)
Rounddummy	0.05***
	(0.009)
Age	0.36***
	(0.003)
Female	0.01**
	(0.005)
Scheduled tribe	-0.08***
	(0.013)
Scheduled caste	-0.05***
	(0.013)
Other backward class	-0.01
	(0.012)
School - Private aided	-0.16***
	(0.018)
School - Private un-aided	-0.32***
	(0.020)
EducFree - Yes	-0.23***
	(0.015)
Constant	3.48***
	(0.044)
Observations	186358
Adjusted R^2	0.739

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Difference-in-Difference Analysis

	(1) Years of education attained
Interaction	0.29***
	(0.078)
Scholarship_ratio	-0.10*
	(0.049)
Rounddummy	0.00
	(0.015)
Age	0.36***
	(0.003)
Female	0.01**
	(0.005)
Scheduled tribe	-0.08***
	(0.014)
Scheduled caste	-0.05***
	(0.013)
Other backward class	-0.01
	(0.011)
School Type - Private aided	-0.16***
	(0.018)
School Type - Private un-aided	-0.32***
	(0.020)
EducFree - Yes	-0.23***
	(0.015)
Constant	3.52***
	(0.043)
Observations	186358
Adjusted R^2	0.740

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7 References

1. *Bagde, Surendrakumar, et al. "Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India." American Economic Review, vol. 106, no. 6, 2016, pp. 1495–1521., https://doi.org/10.1257/aer.20140783.*

2. *Cassan, Guilhem. "Affirmative Action, Education and Gender: Evidence from India." Journal of Development Economics , vol. 136, 2019, pp. 51–70.*

3. *Hanna, Rema N, and Leigh L Linden. "Discrimination in Grading." American Economic Journal: Economic Policy, vol. 4, no. 4, 2012, pp. 146–168., https://doi.org/10.1257/pol.4.4.*

4. *Lowe, Matt. "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration." American Economic Review, vol. 111, no. 6, 2021, pp. 1807–1844., https://doi.org/10.1257/aer.20191780.*

5. *Ray, Tridip, et al. "Whose Education Matters? an Analysis of Inter Caste Marriages in India." Journal of Economic Behavior amp; Organization, vol. 176, 2020, pp. 619–633., https://doi.org/10.1016/j.jebo.2020.02.011.*

6. *Tiwari, Sandeep Kumar, et al. "Caste-Class Association and School Participation in Uttar Pradesh, India: Evidence from NSSO Data." International Journal of Educational Development, vol. 79, 2020, p. 102292., https://doi.org/10.1016/j.ijedudev.2020.102292.*

7. *2015.71655.Annihilation-Of-Caste-With-A-Reply-To-Mhatma-Gandhi. (n.d.).*

8. *Bidner, C., Eswaran, M. (2014). A Gender-Based Theory of the Origin of the Caste System of India *.*

9. *Goel, S., Husain, Z. (2018). Gender, caste, and education in India: A cohort-wise study of drop-out from schools. Research in Social Stratification and Mobility, 58, 54–68. https://doi.org/10.1016/j.rssm.2018.10.002*

10. *Kingdon, G. G. (2007). The progress of school education in India. Oxford Review of Economic Policy, 23(2), 168–195. https://doi.org/10.1093/icb/grm015*

11. *Muralidharan, K. (n.d.). Reforming the Indian School Education System.*

12. *Muralidharan, K., Das, J., Holla, A., Mohpal, A. (2017). The fiscal cost of weak governance: Evidence from teacher absence in India. Journal of Public Economics, 145, 116–135. https://doi.org/10.1016/j.jpubeco.2016.11.005*