A century of unequal opportunity in India: Concentration of educational access and investments

Abhijit Tagade
Master APE
EHESS and Paris School of Economics

Advisor:
Professor Thomas Piketty
EHESS and Paris School of Economics

Referee:
Professor Abhijit Banerjee
Massachusetts Institute of Technology

JEL Codes: I24, N35, D63, H52.

Acknowledgements

This thesis has come to fruition due to the support and patience of my advisor, Thomas Piketty. I owe him immensely. A special thanks to Abhijit Banerjee for agreeing to referee the thesis. Without their body of work, this thesis would not have been possible. I would to like to thank Lucas Chancel for his continuous motivation and support. I thank Olivier Compte for his mentorship, time and wisdom throughout my studies at PSE. My heartfelt gratitude goes to the history group, especially Jerome Bourdieu, Lionel Kesztenbaum, and Gilles Postel-Vinay for introducing me to the beauty of historical research. It is rare to find such a warm and supportive group. Lastly, I thank NIEPA for providing access to DISE data.

I thank my friends and colleagues Resuf Ahmed, Andrei Kalk, Thomas Pastore, Manpreet Singh and Aymeric Vie for highly productive discussions and making the two years enjoyable. The second floor family of Maison de l'Inde and close ones have been a strong support system throughout. Indeed, my journey so far would not have been possible without my parents. Words will not do justice to express my gratitude for all that they have been and done for me—*thank you*.

Abstract

Using administrative and survey data, we study the unequal evolution of education in India from the colonial period to the present. There are two parts to this analysis: access and investments.

First, we find two key factors governing access: minority status and family income. The gap between the top 1 percent and bottom 10 percent is around 10 years of schooling, and gains in schooling have been proportional to family income. The gap between minorities (i.e. Muslims, SCs, STs and OBCs) is about 2-3 years of schooling with respect to non-minorities (Brahmins, Jains, Christians, Sikhs and Parsis) for every level of income. Enrollment data by level and attainment of schooling show identical patterns for the two communities separated by a century. We posit that systematic differences in educational access between the two may play a role in shaping views on opportunity and consequently ideology.

Second, we estimate that nearly 30 to 35 percent of all investments in education went to the top 5 and bottom 50 percent enrollees in the colonial period between 1870 and 1915. That said, inequality of investments declined gradually due to surge among first generation primary enrollees. Since 1980, inequality of investments has increased with the share of bottom 50 dipping to about 30 percent from about 38 percent between 1986 and 2014. The top 10 percent enrollees equal the bottom 50 in the most recent estimates, an equality last seen for birth cohorts from 1870s. Through state disaggregations, we find that the role of public contribution to reducing inequality of investments has increased over time. Hence, proposals to increase privatization to expand access for the bottom groups may exacerbate inequality.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

A society that rewards luck cannot be free. Inequality of opportunity reflects the role luck rather than work plays in determining the distribution of outcomes. Opportunity is thus the cornerstone of freedom, and education forms the basis of opportunity and mobility by imparting knowledge, skills and capabilities to live a life based on choice rather than fate. (Sen, 1985) In the absence of an intervening force, unequal opportunity can reinforce itself through the systematic inequality of outcomes. In the long run, such inequality can cleave communities, and cleaved communities may develop a mutual antagonism and divergence of preferences that may lead to adverse political outcomes. Moreover, unequal opportunity may inhibit economic growth by stifling competition, innovation and economic choice. In short, the degree of equality of opportunity matters as a means and an end in itself, and the nature of its historical evolution may help explain contemporary cleavages between communities.

Despite its relevance, interest in inequality of opportunity has only been recent in economics, and most evidence has so far focussed on the US. Little evidence has emerged from India a la Chetty et al. (2014) on the subject. Being a society by and large segregated by the two key ingredients of opportunity—education and occupation—i.e. along the lines of caste and religion, a colonial past that reinforced social hierarchies, and governed by secular, democratic constitution that mandates education as a fundamental human right today, India presents an incredible site to understand the dynamics of opportunity in the long run. How unequally has education evolved in the long run? Do the social boundaries continue to remain relevant today, or has the increased role of markets transformed the caste segregations into class segregations? Has India been investing in too few at the expense of the rest?

As an attempt to respond to some of these questions, this thesis produces evidence on distributional educational expansion and spending in India from the colonial period to the present. There are two parts to this thesis: access and investments. In the first part, we generate facts about the historical evolution of differences in education between major communities in India. We generate estimates on the long run supply of school, differences in school access and the evolution in mobility by family incomes. In the second part, we produce novel estimates of the historical evolution of concentration of educational investments for colonial period and the modern counterpart separated by a century.

This quantitative narrative is based on a number of survey and administrative data sources, including the National Sample Surveys, Indian Human Development Surveys, British Indian

Censuses, statistical tables of the Progress of Education in India reports and District Information System for Education unit level data. For estimation of general statistics such as population and its growth, public expenditure on education, etc., we have to a large extent relied on public databases such as the UN Population, Clio-infra, World Bank or Ministry of Human Resource Development. The data presented are either at the level of population or representative at the national level. To best of our knowledge, the combination of data sources used and evidence produced in this study are novel.

We find a number of results. First, from less than a year of schooling on an average for birth cohorts from 1880s to completing elementary school today, education rapidly expanded over the last 150 years. This expansion was unequal between men and women, urban and rural areas, and between the forward castes and the rest, aligning with the conventional story. A large part of inequality of opportunity is explained by income differences than social differences. The top 1 percent income households exceed the bottom fractiles by over 10 years of schooling and over 40 times in college completion. Income thresholds may play a key role in determining college completion given bottom 80 percent have more or less equal chance conditional on completing secondary school. Decompositions of access by caste and religion reveal systematic differences in educational access over about 2 years between underprivileged minorities[1] (UPM) and privileged minorities[2] (PM) for every level of family income in any given year, including the top 1 percent. Moreover, the PMs and UPMs replicate patterns of attrition and persistence in school separated by almost a century. Including current enrollment data, the latter continue to remain the only underrepresented group today in every state, whereas the former the only ones overrepresented. Our evidence for long run persistent differences in educational access between PMs and UPMs that transcends class may explain why the two may form divergent views of opportunity.

In the second part of our analysis, we construct the first known estimates of the top share of educational investments in India. For birth cohorts between 1870 and 1915, the share of investments in the top 1 percent has been around 40 percent when accounting for full population and about 10 percent when accounting for only those individuals who attended school. Among just those attending school, the share of investments in top 5 percent equalled that of bottom 50 percent. Given that opportunity of education was largely utilized by the privileged minorities, namely Christians, Brahmins and Parsis, we find that the concentration may have lead to a sizable historical advantage. Between 1980s and 2010s, concentration of total investments grew; the share of investments among the bottom 50 percent educated has declined by about a fourth from 37 percent to 29 percent, and the share of top 10 percent grew from 24 to 29 percent. This equality among the top and bottom groups held last true for cohorts from 1870s. Moreover, decompositions of total and private investments demonstrates that the role of public funds in reducing inequality has increased over time. Therefore, the

[1]Muslims, Scheduled Castes, Scheduled Tribes and Other Backward Castes.
[2]Brahmins, forward castes, Jains, Parsis and Christians.

need for public investments to ensure opportunity remains equitable has increased, thereby undermining the need to move towards greater privatization to ensure equal opportunity.

The findings lead to a mixed policy perspective. First, opportunity gaps between caste and religion separated communities have historically remained substantial, systematic and persistent, and at the same time, the role of family income in determining access is undeniable. Therefore, although caste based affirmative action policies remain relevant, ignoring income inequality in the long run may generate new cleavages based on class identity. Education of Muslims needs to be prioritized in a manner similar to SCs and STs, and our findings support the aims of Sachar (2006) Committee. Targeted financing to benefit Muslims, SCs and STs, and identifying the reasons for their dropout in school is essential to bridge political and social gaps between communities.

If the era of occupational segregation along community lines is being replaced by a competition for a the same pool of resources in labor market due to technological change and increasing returns to skills, then the underprivileged minorities may be the most adversely affected due to their historical disadvantage in accruing the skills, signals and networks necessary to flourish. Especially if the new competition pool is one historically dominated by the privileged minorities, then an increasing share of labor force by the educated members of underprivileged groups may generate animosity. Because markets are likely to rewards skills in the coming decades, it is absolutely essential to ensure that educational gaps between communities are bridged at the bottom of the income distribution.

The thesis is organized as follows. In this chapter, we first introduce context and an overview of literature. Second, we discuss our goals, methods and the data needed to achieve them. In the second chapter, we discuss access. We begin with estimates of long run schooling, followed by income, gender and historical caste dynamics. In the third chapter, we discuss our method to estimate concentration of educational investment and then estimate values for colonial and the recent thirty year period. We conclude finally with limitations, potential for future work and policy implications.

1.1. Background and review of literature

> Education, then, beyond all other devices of origin, is a great equalizer of the conditions of people.
>
> — Horace Mann

Education has been a key engine for global development, and its role in income formation has been well documented. (Easterlin, 1981) More education leads to higher productivity and higher productivity translates to increases in incomes. (Becker, 1962, 2009) Therefore, the distribution of education plays an important role in shaping the labor income distribution in a market economy. (Mincer, 1958) Naturally, education is a key channel for economic mobility of the individual and household, and therefore a systematic difference of access serves as an indicator for the level of unequal opportunity.

Educational access has been afforded to few in India over the course of its history. Aggregate data from Barro and Lee suggest that despite an early start to schooling in mid-nineteenth century, India heavily focussed on tracking to college rather than an wiser expansion in elementary or primary schooling.[3] Despite a greatly lower population shares of enrollment in primary school in comparison to countries such as Japan and France, secondary school enrollment shares in colonial India exceeded them.[4] (see Chaudhary, 2009) Part of the explanation may go back to the educational policy during early colonial rule that expressly focussed on English education of a few natives. Despite a progressive move towards vernacular expansion for the common denominator since, mass schooling neither expanded nor were the recommendations towards compulsory schooling picked up by the British officials as late as early twentieth century citing a lack of demand. [5] Moreover, some evidence suggests that elites exerted influence against mass expansion. (see Basu, 1974; Nurullah and Naik, 1943; Whitehead, 2005) Indeed, not being a major industrial base further skewed incentives against the need for mass schooling.

Consequently, enrollment concentrated among few groups including Jains, Christians, Parsis and Brahmins; minorities such as non-forward castes (including what today are known as Scheduled Tribes, Scheduled Castes and Other Backward Castes) and Muslims had much lower enrollment regardless of provinces in India. (see British India Census 1901, 1911, and Borooah and Iyer, 2005) Occupational differences between caste-religious groups and opportunity costs played a key role in household decisions to educate, whereas on the supply side, schools were likely to be established among communities with a higher prevalence of high skilled occupations. Higher caste-religious diversity led to lower levels of education (due to cost of coordination). (see Chaudhary, 2009; Alesina et al., 1999) While simple correlations demonstrate that concentration of Muslim population was inversely correlated to their enrollment, a channel that may explain Muslim enrollment is the period and place Muslim political decline, i.e. long run historical institutions. (Chaudhary and Rubin, 2011) Notwithstanding the reasons for differences in educational distribution by level or community, evidence points towards a concentration on the education of elites at the expense of mass expansion combined with caste-separated access and occupational structures; indeed, the seeds to unequal distribution of education were sowed in early Indian history, and the overall patterns did not change during the colonial period.[6]

The post Independence period (since 1947) witnessed large increases in aggregate enrollment. (see Kingdon, 2007) However, not much changed in the structure of equal opportunity

[3] See Fulford (2014) for decomposition by birth cohort.

[4] See Green (1997, 2013) for detailed historical account of the emergence of education in the West and East.

[5] See Woods Despatch (1854) and Chaudhary (2009) for a detailed review on history on Indian educational institutional and policy in colonial period. Basu (1867) provides an overview of the education in early colonial period of East India Company's rule.

[6] See Chaudhary (2015) for a detailed history of how caste and education coevolved during colonial India.

post Independence either. The richer and more educated echelons of Muslim society migrated to the newly formed Pakistan and Bangladesh during the partition. (Engineer, 2001) Among those who remained, Muslims, Scheduled Castes and Scheduled Tribes began with similar, low starting points in comparison to forward castes, and gaps in enrollment by level of schooling did not close as recently as 2000. (Desai and Kulkarni, 2008) Unfortunately, although affirmative action policies have gained weight since the commission of Mandal (1971) and caste separated inequalities have reduced, little has been done to eliminate the systematic differences of access by religion despite evidence and recommendations. (see Sachar, 2006) That said, not all states have fared equally poorly in bridging gaps. Decompositions over time and states suggest that left leaning states such as Kerala demonstrate higher levels of opportunity in comparison to right leaning states such as Gujarat and Rajasthan. (see Asadullah and Yalonetzky, 2012) A large share of these differences are governed by differences in primary schooling since higher education formed a negligible share of total education until the turn of the millennium. Given the role primary education plays in increasing social returns and a broad range of benefits including enrichment of public discourse and improved political accountability, these differences in opportunity are likely to cause variations in basic development of states beyond income growth. (Dreze and Sen, 1999)

While caste based occupational differences and occupational inheritance shape the differences between communities in their schooling, in the absence of a system of free and compulsory education, a second key channel that determines educational attainment is the investment decision of the household. Household decisions are constrained by the ability to incur direct (tuition, books, transportation, etc.) and bear opportunity costs (lost years of earning) of schooling. Consequently, these constraints inhibit intergenerational mobility. A large body of emerging literature on contexts around the world provides consensus that parental income and education play a determining role in educational access and future incomes of children despite early educational expansion and high average levels of educational attainment among the population.[7] Rising income and wealth inequality in India on aggregate and along caste lines exacerbate existing unequal opportunities. (Bharti, 2018; Chancel and Piketty, 2019) Indeed, economic inequality translates broadly to decreasing intergenerational mobility, underlying the need to understand and improve human capital for the bottom groups. (see Corak, 2013; Krueger, 2012) Unfortunately, equivalent literature on mobility in India is wanting in the Indian context, whose relevance increases given both caste and moneyed elites have continued to have a historically privileged access.

It is in such situations that the role of public spending and expansion becomes salient to improve access for the bottom groups. A volume of evidence indicates that education yields among the highest returns to investments (see Psacharopoulos, 1981, 1985, 1989; Psacharopoulos and Patrinos, 2018; Patrinos and Psacharopoulos, 2002). Moreover, theory

[7] See (Chetty et al., 2014) and Lee and Solon (2009) for evidence on the US; Deutscher and Mazumder (2019) for Australia; (Bratberg et al., 2017) and Brunori et al. (2013) for international comparisons.

and empirics align that human capital plays an important role in improving growth and development of nations through improved quality of labor and innovation. (Mankiw, Romer, and Weil, 1992; Lucas Jr, 2015; Barro, 2001) Unfortunately, educational spending in India has historically not been a policy priority. Repeated calls for increasing investments in education (for instance the recommendations to allocate over 6 percent of GDP to public schooling by the National Education Commission; see Kothari, 1970) have gone unheard, and the share in national spending has remained virtually unchanged in the long run. (Tilak, 1997a, 2006) Moreover, the starting point for educational investments have been particularly low for the country in comparison to the rest of the world since the colonial period (Chaudhary and Garg, 2015; Davis, Davis, Huttenback, and Davis, 1986; Lindert, 2004).

Among enrollees who benefit from the (low) spending that India undertakes, the quality of educational access has been abysmally low. Grade 5 students are unable to perform basic arithmetic and reading expected at the primary level[8], and PISA 2012 scores ranked India globally and worse than comparable developing countries. College education similarly has been known to produce unemployable graduates en masse.[9] On the other hand, select schools and colleges continue to have a legacy of producing high quality graduates. Rising privatization of the educational sector since the economic liberalization of 1990s has increased the challenge of equitable access further by providing high quality private schools for the middle class and the rich households. (Kingdon, 2020) While we do not contest improvement of quality and spending at the top end, we argue that given limited resources, the need to pull the bottom up has never been more given that their levels of schooling have remained historically low. Lower quality of schooling can further undermine growth.(Hanushek and Kimko, 2000; Hanushek and Woessmann, 2010) An argument is frequently made that market competition and low cost private schools for the poor may boost access, particularly since they may be more cost effective than public schools. (French and Kingdon, 2010) However, they remain inaccessible for the bottom deciles and may not be an effective way to improve mobility. (Harma, 2011) Although higher expenditures do not necessarily translate to rising quality (Hanushek, 1981), given the large dispersion of quality and low absolute levels of schooling at the bottom, we expect that a good share of inequality of opportunity emerges from unequal financing. Unfortunately, while some analyses have examined variations in cost of education in India, there is little evidence on the inequality of both public and private investments outside of the US and Europe. (see Bonneau, 2019; Gounden, 1967; Tilak, 1988) Therefore, a compelling argument can be made on the usefulness of understanding how much of private and total financing benefits the top 10 percent versus say the bottom 50.

Taking stock, unequal opportunity finds early roots in Indian history. Caste and religious separated access to education were exacerbated by the demand of elites to concentrate

[8]See Pratham's ASER reports since 2005 for consistent evidence.

[9]See India Today (2019): https://www.indiatoday.in/education-today/news/story/over-80-indian-engineers-are-unemployable-lack-new-age-technology-skills-report-1483222-2019-03-21

investments in tracking education from primary schooling to college over financing of mass expansion; the lack of a solid industrial base in colonial India contributed further. Since Independence in 1947, these inequalities have continued along caste and gender lines, not all of which are explained by occupational differences and labor market bias.[10] Indicatively, inequality of opportunity tends to be lesser among left states than right states. However, long run evidence on supply of schooling and persistence of structural inequity is lacking. In particular, persistence of cleavages by caste-religious groups deserve greater scrutiny in combination with the role of family income in shaping access. These are especially important since "ethnic capital" may continue to perpetuate inequality. (Borjas, 1992) The role of public spending in improving access has been recoded since the colonial period. (Chaudhary, 2010) Therefore, the recent surge in private spending in combination with elite concentration begs the question of whether India continues to over-invest in a few at the expense of the rest.

With recent developments in skill biased technological change, premiums on high skilled education and generalizable skills have increased. (Autor, Levy, and Murnane, 2003; Goldin and Katz, 2009) Although the penetration of automation in tasks related to manufacturing and primary sectors has been low in India given low labor costs, it may not remain so forever given their substantively high productivity to perform a wider range of tasks cost-effectively. (Aghion, Jones, and Jones, 2017) Rapid increases in human capital is essential for communities in India to survive globalized competition. Communities with traditional occupations and low levels of schooling are the most vulnerable to technological and globalization shocks since skill-substituting meaningful jobs may simply not continue to exist at low levels of education. From a broader state building perspective, the role of high skills in contributing to development has been well recorded in India since early days. (Castello-Climent et al., 2018) Therefore, the need for evidence on sources of persistent inequality of opportunity are central not only for the purpose of ameliorating it, but also to inform policy decisions on ensuring that the historical cleavages do not impede the long run growth and development of India.

1.2. Scope and method

> Finding the question is often more important than finding the answer.
> — Tukey (1980)

Our review of literature highlights key gaps in evidence on inequality of opportunity. In particular, a representative, quantitative narrative is lacking on the social bases of historical divergences in educational access and the evolution of concentration of educational investments. Rather than writing this section once we have some evidence as if we were hoping to find the relevant phenomena all along, we confess that this thesis is a Bayesian quest to raise

[10] See Kingdon and Theopold (2006); Fasih et al. (2012); Fulford (2014) for variations in returns to schooling.

further questions by presenting new evidence on the anatomy of unequal opportunity from the colonial period to the present.

Some directions are clear. On aggregate, what does the long run school school supply in India look like? How does parental income affect access to education? Are the average gaps of access between minorities and non-minorities substituted or complemented by parental income? How are these differences between communities governed by a state's inclination to an ideology? With increased privatization, has the country increasingly been investing too much in the human capital of too few? What similarities persist in this concentration from a impoverished colonial period to one where education became a fundamental human right? Do differences in state ideologies matter in investments as they do in access?

This thesis in economic history is a collection of evidence on these questions using the tools of exploratory data analysis. (see Tukey, 1977) A large part of our work has been the construction of a historical and contemporary data series and only a small fraction of the data compiled are reflected in the evidence presented. At large, what are *not* of interest to us are the particular point estimates (of say differences in years of schooling between two adjacent percentiles of household income), but the key trends underlying data (such as systematic differences in the nature of increase in years of schooling over a range of income percentiles between two groups). The latter *will* constitute a large part of our inference.

Our analysis is constrained by availability of quality data. Our earliest time series of educational access go back to birth cohorts of 1890s and the most recent cross sectional evidence covers up to 2016. Estimates of the concentration of educational investments go back to birth cohorts of 1870s, and the most recent cross sectional ones date 2014. Despite availability of institutional and enrollment data dating as back as 1850s, their representativeness may be constrained by the administrative reach of the British rule in India. In the next section, we describe the various data sources used towards the analysis.

1.3. Description of data and constructions

We combine a number of survey and administrative sources in this thesis. For colonial India, we utilize British Indian censuses and the statistical tables of Progress of Education in India (PEII) documents which report population level statistics for British India. For post independence period, we use the employment-unemployment, education and consumption-expenditure rounds of National Sample Surveys (NSS) beginning 1983 to 2014, the ICPSR Indian Human Development Surveys (IHDS) of 2005 and 2011, national statistics compiled by the Ministry of Human Resource Development (and what was previously Ministry of Education and Social Welfare) and level unit data of District Information System of Education (DISE) organized by the National Institute for Educational Planning and Administration (NIEPA) for the years 2005 to 2018. The NSS and IHDS report representative data at the state and national level, whereas DISE represents nearly population level school data.

1.3.1. Progress of Education in India and British Indian censuses. Although the British Indian Censuses commenced being collected in 1881, PEII reports and its variations (that we will collectively refer to as PEII) were being compiled by the Imperial administration since the 1860s; this was also the period when British administration was being established over native states. Every publication of the PEII was split by two volumes: the first clarified definitions, described the problems and opportunities of educational situation of then India, the policy plans of the British administration and their outlook for coming years; the second volume provided the statistical appendices to the first volume. The latter are utilized for the purpose of understanding trends in educational enrollment, and supply of schools and colleges which we consequently use to estimate the concentration of educational investments for the purposes of this thesis. The data increasingly compiled over increasing geographies, covering much of what today are known as India, Pakistan, Bangladesh and parts of Myanmar (then Burma), with some variation as further territories came under the purview of the Imperial administration for statistical purposes. The PEII reports were produced quinquennially, and although initial reports provided data at an annual level by provinces, later variations around the turn of nineteenth century deferred to quinquennial numbers.

To access PEII, we utilize the internet archive website *archive.org*, the South Asian digital archives of The University of Chicago and the digital library of Gokhale Institute of Politics and Economics. For the share of years where data are not digitized, we manually record observations from the archival documents. There are some inconsistencies in the colonial Indian PEII reports over the years, particularly the nominal values of expenditure in education depending on the year of report (for instance, the national expenditures do not tally with the total in the 1917 version of the statistical appendix when accounting by source of expenditures) and currencies used for expenditure (some years in nineteenth century use the pound, others use rupee), among what are noticed immediately. Some of the inconsistencies are attributable to differences in accounting, but we check for internal consistency over the years to produce a reliable time series. Furthermore, there are differences in the content of statistics reported over the years in PEII. Select years report enrollment data by caste (including separation by Brahmins and non-Brahmins), some report details of higher education by degree types, and others do not. We pick the best available evidence from available reports to get detailed disaggregates for an accurate estimation of the concentration of investments in education. Due to constraints of space and time, and the intensity of work involved in the archival process, the data are compiled until 1930.

The British Indian censuses were utilized to verify population estimates of various schooled groups and total number of individuals. We also use Mukherjee's (1976) estimates for age wise population by decade for construction of concentration of educational investments that require the number of unschooled individuals.

Schooling in colonial India was organized into primary, middle and high stages for conventional schooling, which included vernacular and English medium schools. Schools in the standardized system were classified as primary, middle or secondary, although some middle and secondary schools also included primary grade students. There were also parallel systems of "training", "private" or "special" schools which catered to numerous groups, including religious instruction, vocational training for professions, trades and crafts and sometimes for education of the nobility. These parallel schools have been broadly classified as "special" for our purposes which broadly fall in line with the classification by PEII reports. The reports ensure to separate higher level training schools, such as for law, medicine, engineering, etc. into professional college programs to avoid confusion. Schools were governed by both provincial administrations directly as well as managed by private individuals but funded by public funds. The institutional and enrollment data in PEII encompassed their population by management and funding type. For our purposes, we utilize the totals. College education was separated by arts and professional streams, with post-graduate often at the university level. We utilize the total of arts and professional colleges for our data for each year.

The definitions of primary, secondary and college education have not varied greatly over time, although some variations existed between provinces. For instance, primary education has consistently been defined as somewhere between 4 and 6 years of schooling, middle education between 2 and 3 years and high stage between 2 and 3 years, totaling the full period of schooling between 9 and 11 years. (see PEII Volume I reports) In line with broad definitions in Volume I of the reports over a number of years, we standardize these definitions at the national level for consistency at our end with 5 years of primary schooling, 3 years for middle stage and 3 years for the high stage until matriculation. Bachelors degrees lasted roughly two years, although a handful lasted for three years in the twentieth century. We standardize all post-secondary schooling to two years for our purposes, which also helps balance the extended estimate for schooling to provide a conservative of inequality. Given a high per student expenditure in special schools and their often professional nature, we categorize them as parallel to secondary education but lasting for about four years.

A briefer analog of the PEII reports continued until 1976 which were produced by the Government of India after the Independence in 1947. Such quality data as the PEII series has not been available since, thereby constraining a comparable analysis severely for the recent decades. Although we utilize the aggregate statistics on enrollment, public expenditure and institutional expansion provided by the MHRD in recent years, they do not provide the level of transparency and disaggregated information that PEII did.[11]

1.3.2. National Sample Surveys. The digitized versions of National Sample Surveys since 1983 present the best available representative data series beginning in 1983 thereby

[11]The lack of data transparency on education that we seek commenced after the period of political turmoil at the central government, namely the Emergency during 1975-77.

bridging the data gap for recent decades. We mobilize the "Consumption-expenditure", "Employment and unemployment" and "Education" rounds of the NSS from 1983 to 2014 to estimate enrollment, educational expenditure and household consumption differentiated by caste, age demographics and states. Due to availability of household level information on both education and household expenditure we use the employment-unemployment rounds (accessed from IPUMS) for estimating unequal access. The categorical variable on education "educing" is defined by illiteracy, below primary, primary, middle, secondary, (in recent rounds) higher secondary, and college or post-graduate. We construct years of schooling using this data as 0, 3, 5, 8, 10, 12 and 15 respectively, with the exception for 1983 where we define secondary as 11 years of schooling owing to 11 years until matriculation from the high stage in a number of states prior to the survey. For secondary school completion, we define individuals who have completion at least 10 years of schooling and for college at least 15 years of schooling.

For household consumption data, we construct monthly per capita household expenditures when not directly available as a ratio of total monthly household expenditure and number of household members. Total educational expenditure data are standardized to monthly expenditure adjusted for length of educational program enrolled in that include tuition, other fees, stationery, textbooks, transportation, uniform and private coaching. Not all expenses are incurred by the household, and the poorest receive assistance from the government for each of the private expenditure component, in addition to scholarships, tuition waivers and mid-day meals. Some grants are received in full, some partially, some with subsidy, and others with both. The data for government expenses are not provided in nominal values for each education related expenses. The size of the full grants are estimated for textbooks, uniform, stationery and transport as the average size of the private analog of the assistance program (for instance, full textbook grant is assumed to equal the average private textbook expenditure) per round of survey; the size of partial and subsidy is assumed to equal half, and their combination as three fourth of the average private expense. Since the component wise private investments are reported for roughly 12 months, we divide the the public contribution by the same ratio to have comparable values. We fully realize that these represent rough estimates, especially given that private expenses are likely to be skewed towards the higher end and are not identical for every level of education or by location (given price indices vary), but they still provide meaningful starting points for public contribution in education at the individual level. Education is reported to be free for a large although decreased share of enrollees over time: from nearly 75 percent in 1986 to about 48 percent in 2014. To estimate the public contribution for free education, we utilize education and year level wise weighted average of tuition expenditure of those who do not receive free education.

To produce comparable expenditure values over the years, we adjust them for CPI inflation at the 2014 level, whose data we obtain from the World Bank. Using recent mid-day

meal expense data of about INR 5.5 per student per day as of 2014[12], we add government expense of about INR 100 per month for individuals receiving a mid-day meal on the basis of 200 working days in a year and 11 functional months (or around 18.2 working days a month) in sync with other expenses. The component wise public contributions add up to the public counterpart for total private expenses per month, and the two together represent the total investment investment in education of an individual per month.

For caste, we rely on the "social group" classification of the NSS into Scheduled Castes (SCs), Scheduled Tribes (STs), Other Backward Castes (OBC) and Others (that include forward castes, and non backward caste religious groups).

1.3.3. Indian Human Development Surveys. Unlike the NSS, the ICPSR IHDS are among the few population representative data sources for India on household and individual data on education, consumption-expenditure and employment with caste-religious disaggregation, including Brahmins and non-Brahmin forward castes. We use one of the only two IHDS surveys undertaken, namely the 2011-12 version, for most recent estimates of unequal opportunity. This survey records the years of schooling of individuals; however, we recode all years of schooling above 15 as 17 as upper limit for post-graduate schooling. Like in NSS, we record secondary school completion as having completed over 10 years of schooling and college as over 15 years of schooling or having a Bachelors degree. Analogous to NSS, we use the default per capita household income available with the data to estimate the family rank in income. As a complement to findings from NSS, we utilize the IHDS to explore caste specific inequality in opportunity.

1.3.4. District Information System of Education. DISE administrative data cover nearly population level data on all schools of India, and are collected by the National Institute for Educational Planning and Administration (NIEPA). We extract the unit level data of all schools in India (granted access after a formal approval process in 2019) including public, private and partnership schools for all years available (2005 to 2018) that cover numerous aspects of the population of schools including principal name, socio-demographic details of students by each grade level, precise location of school at the sub-district level, infrastructure, teachers, funding and establishment. (This process took months minus the time for approval.) We manually extract over 6000 raw data files for this purpose.

Unfortunately, DISE data contains several discrepancies between years: in particular, we find that older years may have less coverage than recent years given that the number of schools established are underestimated for recent years, or that registry in DISE system may have simply taken time. Figure A.0.1 of the Appendix highlights this discrepancy. For the purpose of this thesis, we utilize it to produce a unique series of year of establishment of all surviving schools, and to analyze grade wise persistence of students by caste groups as the modern counterpart to historical data.

[12]See https://news.biharprabha.com/2014/07/mhrd-increases-cooking-cost-under-mid-day-meal-scheme/.

CHAPTER 2

From privilege to a human right: Educational expansion for 150 years

Two epochs, separated by a century, characterize the expansion of education in India. The first was the rise of formal schooling (as opposed to the prevailing indigenous systems of education) in the second half of the nineteenth century colonial India. And the second was a rapid expansion of basic schooling in the second half of twentieth century Independent India. This chapter broadly examines the supply of education in India from colonial period to the present.

Since the establishment of colonial administration in India, the British expanded schooling as a means to both govern the peoples and to evangelize Rationalism.[1] Few went to school in India in the nineteenth century: learning by doing in hereditary trade was the predominant form of human capital accumulation for most in nineteenth century child labor in various forms (including family enterprises) kept children from school, and gendered roles kept most women from schooling. Like in Britain, formal schooling was largely an enterprise of the elites. Consequently, a large share of the population could neither read nor write, and few outside of nobility elites in India could access formal schooling in the nineteenth century. Literacy in 1900 India was around 5 percent. Among those few who had access to education were notably *savarna* men[2], many of whom would go on to serve as intermediaries for the British administration in India, while others worked as professionals such as accountants, doctors, lawyers, engineers and clerks aiding in fulfilling the goals of the English Act of Education of 1835, and some others had traditional training in matters of religion or vocational trades. From 1850 to 1920, India transitioned into the contemporary notion of schooling (henceforth "schooling"). This transition, however, was largely unequal.

Despite a continuing rise in schooling by the early part of 1900s, lack of access to education persisted through much of the twentieth century, including after Independence in 1947. Lack of literacy remained as high as 25 percent until recently. Since the turn of millennium, India achieved nearly universal enrollment in elementary schooling ushered by the

[1] See the Council of India (1835); Macaulay (1835).

[2] Those broadly identified as belonging to *Brahmin*, *Kayastha*, *Kshatriya* and *Vaisya* castes.

Education for All and *Sarva Shiksha Abhiyan* campaigns in the 1990s and 2000s.[3] For a democracy of India's size and scale that began as an impoverished, colonized and caste separated society, the transition to universal access from 1950 to 2020 was monumental. That said, the recent expansion of school access has also not been egalitarian—significant differences have persisted between men and women and privileged and underprivileged minorities in both quality and quantity. Despite the constitutional mandate of compulsory elementary education granted by the Right to Education Act of 2009 in succession to the universalization campaigns, there remain important concerns of equality of opportunity as we learn from the experience of high income countries that universalized access as early as a century ago and yet face significant inequality of opportunity today.

Unequal opportunity between communities through unequal human capital accumulation reproduce social and economic inequalities through labor markets. Therefore, the need to understand the nature of unequal access cannot be understated. Using a series of novel administrative and survey data, we explore in this chapter three key channels through which education evolved through Indian history: the first is long run aggregate schooling, the second is through differences between social groups, and the third is the role of family in shaping access to education.

We discover the following facts through this process. First, schooling boomed in two periods in Indian history: Nehru's premiership and post-liberalization. The lack of continuity in school expansion post-Nehru deserves attention since it accounts for a very large share of lost human capital. Second, through the twentieth century, the average Indian went from having no education to completing elementary schooling. Differences in schooling between men and women, urban and rural areas, and forward and non-forward castes have persisted in the long run, aligning with the usual narrative. Third, we find that family income plays perhaps the most important role in determining school access. The relationship is convex, and the difference between the top 1 and the bottom 10 percent amounts to nearly 10 years of schooling as recently as ten years ago. Inequality is most sharply represented in higher education where even the younger cohorts among the bottom 80 percent have a less than 20 percent chance of completing a college degree, whereas the top 1 have a nearly 80 percent chance. Despite starting from a much higher point, gains since Independence in educational access have been higher the richer the household.

Furthermore, we find that there are two separate sets of communities with near identical distributions of education in each set—both over the course of the century and a given time over the range of family incomes. These sets are the privileged minorities (consisting of Brahmins, Jains, Parsis, Christians and other forward castes), and underprivileged minorities (consisting of Muslims, SCs, STs and to some extent OBCs). The former have a systematic

[3]Furthermore, India ratified a constitutional amendment, Right to Education Act (2009), that characterized education as a fundamental human right in Article 21a, and mandated free and age appropriate access for all children from the age of 6 to 14.

and persistent premium in education at every level of income relative to the latter. Enrollment shares over levels of schooling also show persistent patterns despite being separated by a century. Moreover, Muslims, SCs and STs are also the only groups still underrepresented in educational enrollment across states of India as of today. Although left states fare better than the right states on an average, the differences are smaller in comparison to the variation.

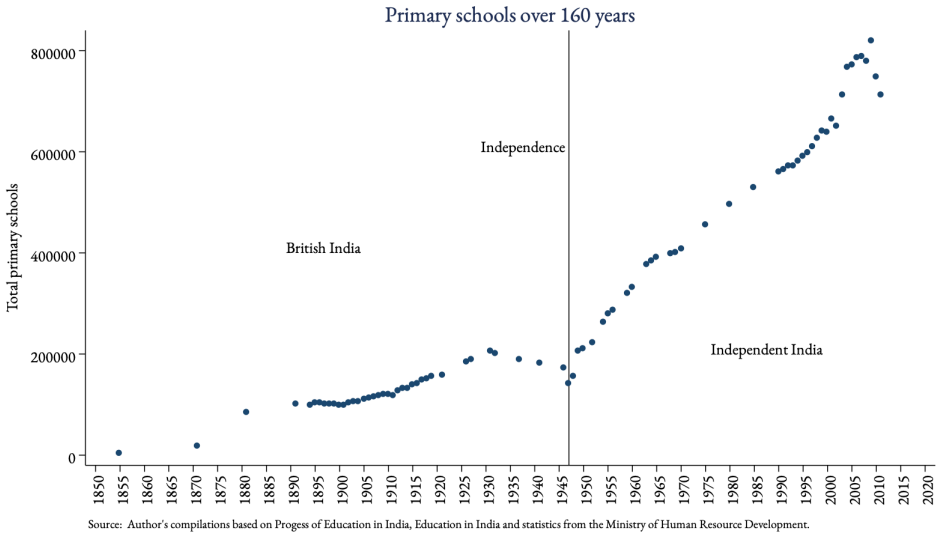
2.1. Two hundred years of expansion

Education greatly expanded over the last two centuries in India; what was once a privilege of the few became, about a decade ago, the human right of all. This section is devoted to exploring big questions on school supply, the evolution of various measures of human capital over the long period and aggregate inequality.

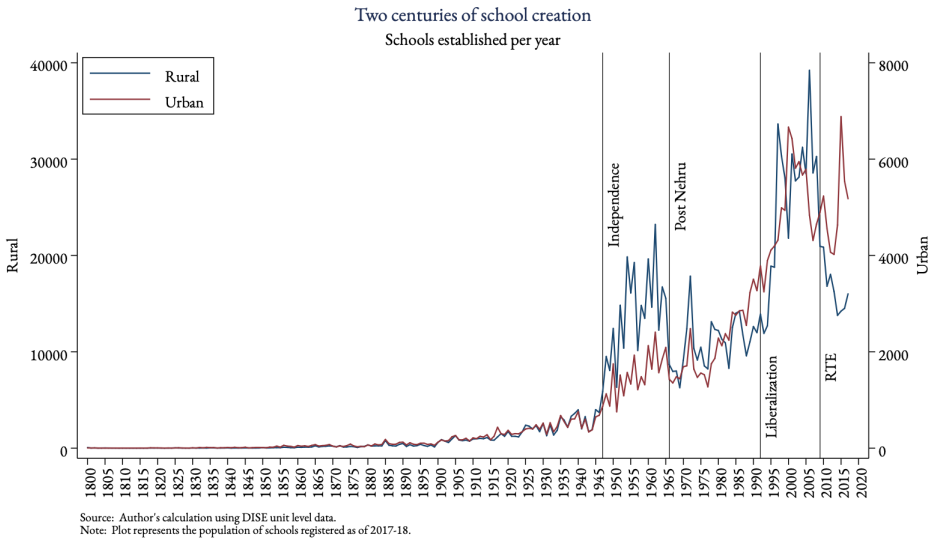
During the nineteenth century, schooling was largely local, vernacular and traditional. Like any other part of the world, it mostly focussed on religious and philosophical education with some combination of localized notions of mathematics and natural sciences. Since most economic activity was based on hereditary trades, education was not a means to mobility but to train on default trades. Consequently, large part of education and training occurred on the job. Schooling first emerged in Europe and the United States over varying years in the 1800s and expanded with the needs of the Industrial Revolution. The Revolution had not consumed the Indian subcontinent the way it did Europe; the expanding rule of the British East India company further brought challenges to its expansion with greater administrative and political instabilities among the local warring states. Therefore unlike Europe, schooling retained its traditional role in the Indian society, and public access to formal training did not gain traction like the working class in Europe did by the end of the century.

The advent of colonial rule introduced formal schooling by importing institutions from Europe and incorporating curricula along the lines of Macaulay (1835). Primary schools became the point of entry, followed by a specialized education at the secondary stage and then greater specialization at the college level. Less than 5 percent had access to primary education in the mid nineteenth century, and most who did were Brahmins in general schooling and Muslims in vocational schooling. Schooling expanded gradually until the time of Indian Independence in 1947. Since then, school supply exploded for a brief fifteen year period during Nehru's premiership, and paced up similarly only after liberalization in the 1990s. Since the turn of the millennium, universalization of elementary schooling increased enrollment rates from about 80 percent to about 100 percent as of the recent decade.

The long run trajectory of basic education is captured in Figure 2.1a. The first part of the figure captures the number of institutions registered as primary schools over the the 160 year period from 1850s to 2010s. To produce this, we use a number of archival and contemporary population level administrative data, including the statistical tables of Progress of Education in India and reports by the Ministries of Statistics and Human Resource Development. Years until 1947 capture British India, including parts of Pakistan, Bangladesh and Burma. Data



(A) School supply in India: 1850 to 2015



(B) School creation in India: 1800 to 2016

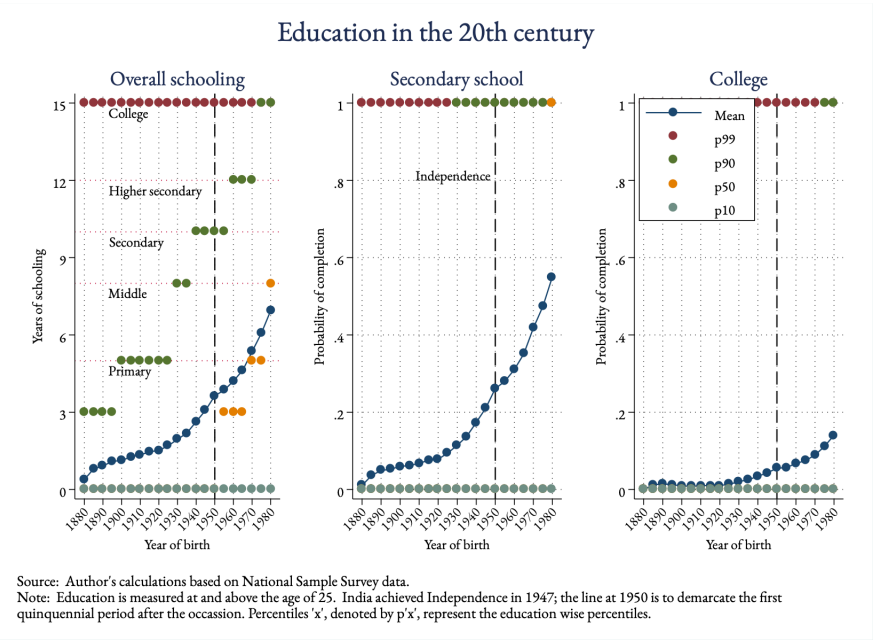
FIGURE 2.1.1. Schooling in the long run

after 1947 capture sovereign India. The fifty years between 1880 to 1930 show a gradual increase in school supply—doubling from around 100,000 to 200,000 institutions with about 2,000 new schools every year. The Nehru years post 1947 mark a sharp rise from by more than doubling within just 15 years as if to match the trend extrapolating 1930 onwards. Unfortunately, the vision for school creation was evidently not matched since until 1990s when markets took over with the introduction of liberalization—up to the Right to Education in 2009.

A much richer description of this image would tell us about the precise number of schools created or destroyed per year. Unfortunately, annualized data are not available for all years. Given that school creation varies substantantially as a policy choice rather than natural rates of growth, we do not believe interpolating yields appropriate rates of school establishment. To address this gap by having a continuous measure, we use the near population level school data from DISE and aggregate unique school registration codes by each year of establishment. Figure 1b presents the number of schools established of all kind—public and private—by year and region as measured in 2017-18. The results are largely consistent with Figure 1a.

There are two major episodes in Independent India where schooling boomed: Nehru years and post liberalization. The impact of these two episodes in bringing literacy and enabling opportunity cannot be understated. Nearly 15,000 schools were created per year for 15 years in rural areas and about 1,700 in urban areas during the first boom, and about 30,000 in rural areas in 5,000 in urban areas during the peak of the second boom. Both rural and urban areas follow more or less identical trends throughout Indian history; rural areas are scaled by a factor of five. The Nehru years are a distinctive period where rural expansion outgrew urban expansion. While understanding the cause the dip in pace of Nehruvian expansion until the 1990s needs further work, the findings conclusively highlight an incalculable and irrecoverable loss of human capital gains during the thirty years since 1965. It is unlikely that the voluminous s expansion was perfectly substituted in equal degree with investments in quality through the full period. The second boom of liberalization was guided by private involvement in schooling and ended sharply with the passage of Right to Education Act in 2009. The reason for the drop is not well understood; some potential explanations include deregistration of schools

Now, by construction, the number of schools by year of establishment reports the number of surviving schools at the time of measurement. Like firms, new schools get created every year, but unlike firms, it is difficult to imagine destruction of schools except in the very long run or due to large shocks. Therefore our data for colonial period under represents schools. The data before 1947 are not representative since the DISE values conflict with administration's reports. A cause for this may be destruction of schools in 1947 or different shares of schools created in different administrative regions not under post 1947 India. Therefore, it



Source: Author's calculations based on National Sample Survey data.
Note: Education is measured at and above the age of 25. India achieved Independence in 1947; the line at 1950 is to demarcate the first quinquennial period after the occassion. Percentiles 'x', denoted by p'x', represent the education wise percentiles.

FIGURE 2.1.2. Evolution of schooling in India by birth cohorts: 1880 to 1980

is instructive to use Figure 1a for an aggregate overview and Figure 1b for trends in creation after independence.

The rising supply of institutions was accompanied with a rise in the country's human capital. Figure 2.1.2 captures the long run trend in years of schooling, the chance of completing secondary school and the chance of having a college degree for a century of birth cohorts since 1880s.[4] What was once a just a tenth of the population having access to schools rose to the average person having completed elementary schooling born in 1980s, which translates to gains of .08 years of schooling per year. The rise has been more or less steady over the course of the century broken only by a small unexplained kink for cohorts born around the time of Independence. The shock likely shifted forward the curve by about 1 year of schooling over a range of five years of birth cohorts.

Our time series suggests that educational access has been substantial. The gap of time between the top 10th and the average person having completed elementary schooling is about half a century, that is two full generations. While the top 10th person had access to high school by a decade before independence, it is only in by liberalization that they completed college on an average. The top 1 percent on the other hand enjoyed college access through the full

[4]To estimate human capital for a birth cohort, we use the age data of individuals reported in rounds of National Sample Surveys since 1983 to trace back the closest quinquennial year of birth and take the weighted average of every measure of education. See data description section for details on measurement of education.

course of 100 years. Our data suggests that educational gains have percolated among the masses at a slow rate.

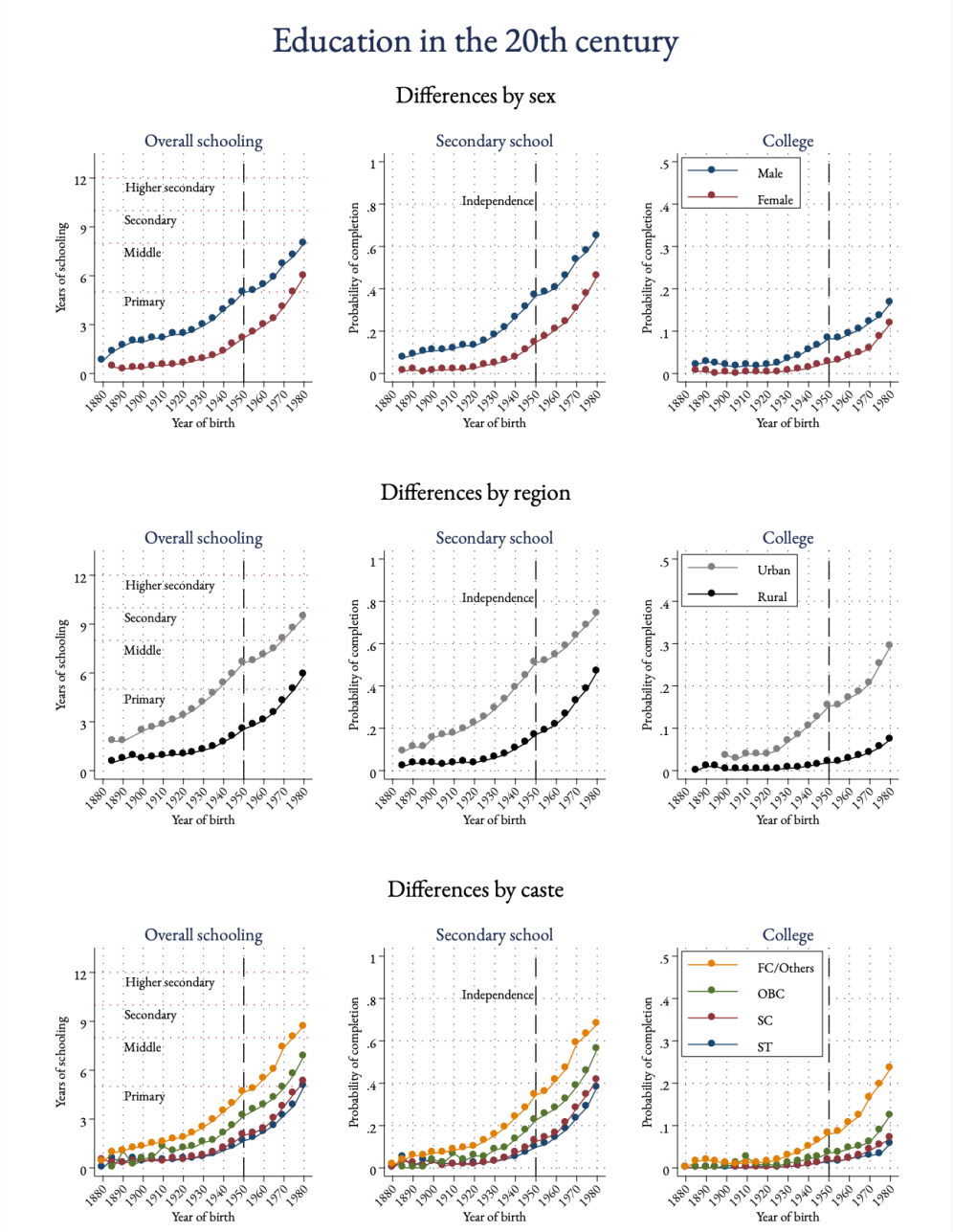
To further examine the sources of variation in inequality of education, we decompose the population by sex, rural-urban region and broad caste categories of the NSS. [5] (see Figure 2.1.3) In line with the usual discourse, we find males, forward castes and urban areas having an advantage over their counterparts. Further, regional variations are demonstrably the largest source of historical inequality in educational in comparison to sex or caste. The average urban resident born in the 1980s had access to post secondary schooling whereas the rural counterpart had not completed middle school. Urbanized areas are likely to attract skilled labor and hence more likely to have schools to produce them further. This is reflected in the sharp gap of over 20 percent in college completion rates. Since men and women are more likely to be somewhat equally distributed between rural and urban areas, the 2-3 year historical gap in years of schooling between the two reflects the social preference for men over women driven by historical gender differences in the labor market. Caste differences may be accounted for slightly by regional variation since castes reflect historical differences in occupations and regions are segregated by occupations. That said, the salience of our result is that unlike sex and region, caste gaps have *widened*—not reduced—over time. Regardless of policies on affirmative action and movements to expand schooling, the "unreserved" caste groups have gained higher access over the "reserved" castes at the margin as recently as the cohorts of 1980s who received their education during the 1980s and 1990s.

The closing gaps between sexes, widening gaps between castes, and opportunity for individuals at the top of educational ladder form the substance of this chapter. In the following section, we zoom in at a given point of time and understand how access to opportunity differs by family income.

2.2. Who has opportunity: The rich take it all

An important question surrounding equality of opportunity is: who has it? Our time series suggests that there have been systematic differences by sex, region and caste. However, differences between groups may be far overshadowed by variations within group, which lead to the former overstating systematic differences in opportunity. Parental income is an important determinant of access to education. (Chetty et al., 2014) However, little is known about distribution of access by parental or household income in the context of India, let alone its evolution over time. This section is dedicated to examining variations within years by the household's income.

[5] Demographic compositions of rural-urban regions evolve and new areas become urbanized over time. Our data captures the values at the time of measurement; while not perfectly precise, they still provide a rough but useful estimate of the size of differences in educational attainment.



Source: Author's calculations based on National Sample Surveys.
Note: Education is measured at and above the age of 25. India achieved Independence in 1947; the line at 1950 is to demarcate the first quinquennial period after the occassion.

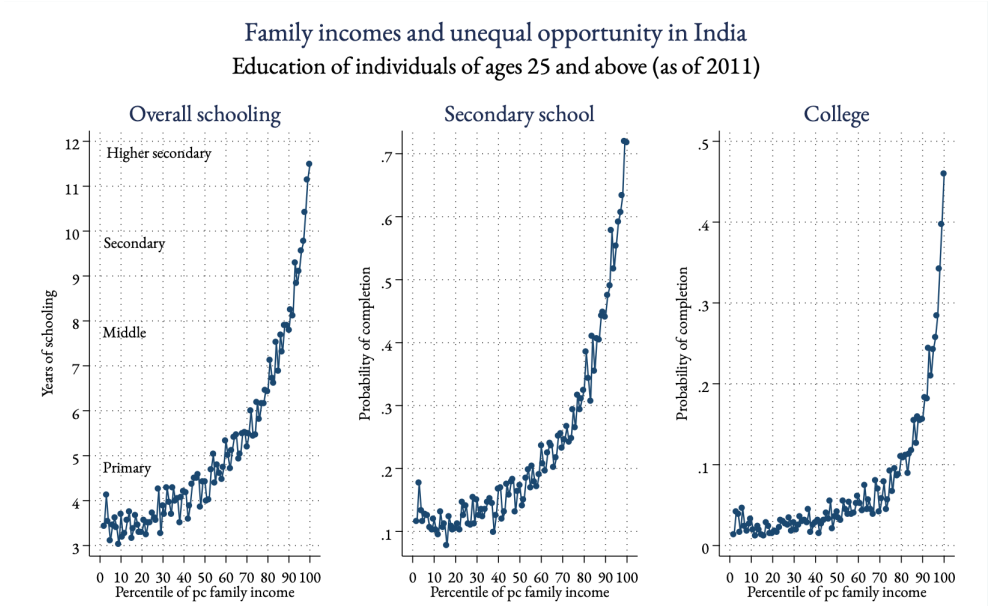
FIGURE 2.1.3. Decomposition of schooling in India over social groups by birth cohorts: 1880 to 1980

Richer households afford more education than the poor. But it is not obvious ex ante to what degree richer households would prefer it over the poor. The degree of preference reveals about household expectations on returns to schooling and the change over time reveals differences in these expectations. Poor households simply cannot afford schooling due to high opportunity costs since the child can assist in parents in their economic activity, which largely lie in the agriculture sector in India. Therefore, to a large extent, levels of schooling among the poor reveal the opportunity cost or budget constraints. The household's effective income is not captured by the total income since family income is correlated to household size. The amount that available for each individual under the roof provides a richer source of comparison of resources. Therefore, our unit of analysis will be per capita household income. Although absolute incomes provide rich insights in assessing how affordable education is for families, we constrain ourselves to relative income since they provide salient insights on how the richest versus the poorest groups fare at any given time. Using representative survey data since 1980s (NSS and ICPSR IHDS), we analyze the evolution of unequal opportunity against percentiles of household incomes. Since NSS data do not capture income but rather household consumption, we utilize the per person consumption values in lieu of income.

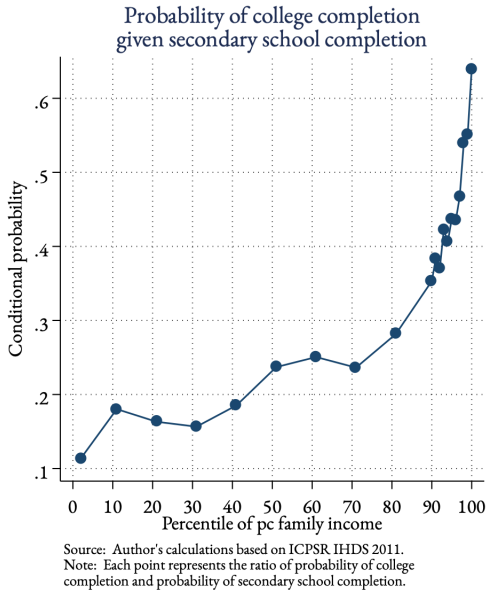
There are three ways through which we decompose educational attainment: overall education measured by mean years of schooling; secondary schooling or matriculation, measured by probability of completion; and college, measured by probability of completion of a Bachelors degree at a minimum.[6] Years of schooling provides a reliable continuous metric of schooling access. Secondary school and college have played a major role historically in separating skilled from unskilled labor. Secondary school signals skills useful in the formal sector for clerical roles and vocational training useful in operation of machines, manufacturing related processes and other semi-skilled professions for which elementary schooling may not be sufficient. Given the substantial opportunity costs in sustaining enrollment until the completion of secondary schooling—which is the first marker of a skilling—it is a useful starting point to assess how does occupational separation occur over generations. Completion of a college degree signals greater skills but also accompanies five years (or 50 percent more) of opportunity and direct costs in excess of secondary schooling. Naturally, only the rich are likely able to access college.

We examine the variation in educational attainment of adults who have completed schooling (age 25 and above) by household incomes per capita (which we will use interchangeably with simply income) in 2011 using ICPSR IHDS in Figures 2.2.1a-b. The bottom 60 percent attain less than primary school on an average, whereas the richest 1 percent nearly complete higher secondary schooling. Over 70 percent of the richest 1 percent complete secondary school and nearly a half possess a college degree—an over 40 times chance than the poorest households. The difference between the richest and the poorest exceeds 8 years of schooling, and almost 60 percent chance of completing high school. The rise in schooling among the

[6]See section on data description for more details.



(A)



(B)

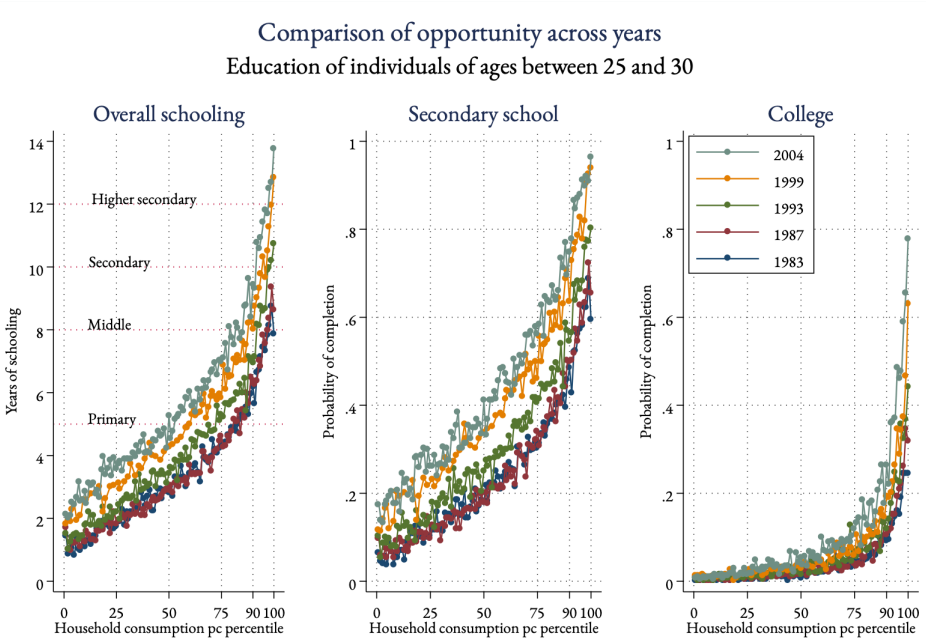
FIGURE 2.2.1. Family income and educational access

top 20 percent is sharp in all three categories and the full distribution follows nearly an exponential curve. Keeping in mind that this includes older demographics when schooling was less prevalent and more expensive to access, the differences in education are substantial by family incomes.

Evidently, the distribution of overall schooling and secondary school follows a similar distribution, whereas college access is far more skewed. A number of individuals who complete secondary school among poorer households may not transition to college simply due to budget constraints. Therefore, the probability of completing college conditional on completing high school provides insights on the ability to afford acquiring higher skills having reached a minimum threshold. We estimate this value by taking the ratio of college completion rates with secondary school completion rates. Figure 2.2.2b represents the percentile-wise conditional probability among individuals of ages 25 and above. The exponential patterns of access are roughly replicated. Less than one in four of the bottom 50 percent completed college having completed secondary school in contrast to the richest 1 percent among which over 60 percent do. The advantage of the top 1 percent exceeds by 30 percent (or a little less than double) that of the average top 10th household. These reflect a remarkable difference of resources allocated to education not only between the poorest and the richest, but among the richest themselves. Part of what may explain the surge among the top 10 is a threshold effect: since incomes are more dispersed at the tail of the income distribution, differences between adjacent percentiles increase exponentially over percentiles. College completion may require access to a minimum pool of funds to sustain studying for five years post secondary completion, and each percentile above 90 is likely to have increased ability to forego earnings for the same period.

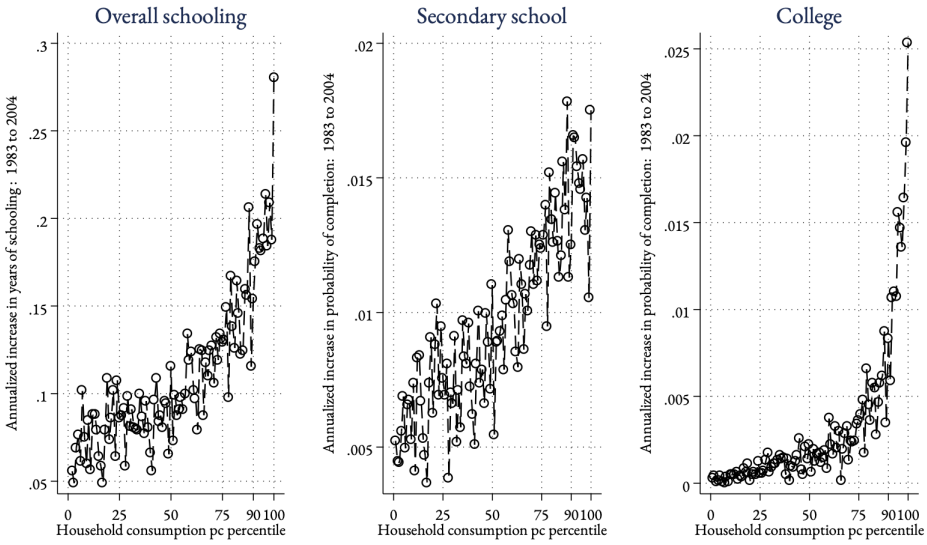
Recent developments in enrollment and supply of schools may suggest that inequality of education for the full population of adults may overstate itself. Younger cohorts tend be better schooled than older cohorts. Moreover, birth cohort wise education for income percentiles at a given time may not represent the evolution of inequality of opportunity since children's income ranks may not map perfectly with parental income ranks. Therefore a salient measure of the evolution of access would be to estimate schooling among younger cohorts by relative incomes over time. We estimate this using NSS from 1983 to 2004 for individuals between ages 25 and 30 at the time of measurement and summarize the results in Figure 5a. The annualized change in overall schooling, and secondary and college completion are summarized in Figure 5b. Due to absence of income data, we utilize consumption wise percentiles and avoid combining 2011 data from ICPSR since the two employ slightly different sampling.

The results are telling: exponential patterns are replicated with higher annualized growth the richer the household. But more importantly, in accompanying unequal access at every point in time, the growth rate varied substantially between households at any given time.



Source: Author's calculations based on National Sample Surveys.
Note: Samples are individuals aged between 25 and 30 years. Household consumption per capita percentiles are estimated for full population.

(A) Evolution of schooling access by family incomes



Source: Author's calculations based on National Sample Surveys.
Note: Samples are individuals aged between 25 and 30 years. Household consumption per capita percentiles are estimated for full population.

(B) Gains in education by family income

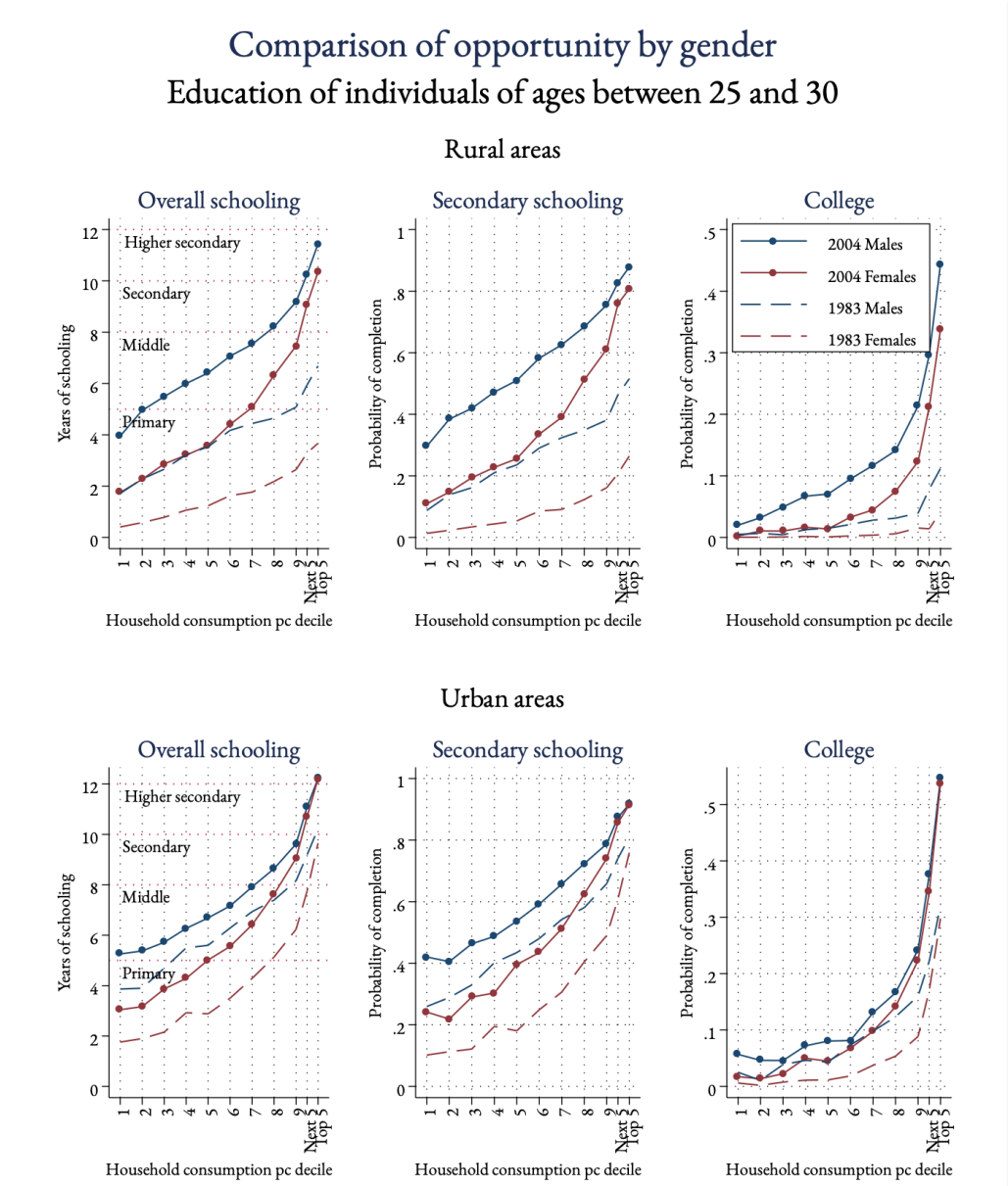
FIGURE 2.2.2. Family incomes and educational access: evolution

The bottom half grew at 0.1 years of schooling annually in comparison to 0.2 years for the top 10 percent after discounting the top 1 percent for sampling reasons. Per year increase in secondary school completion rates differ by about constant 0.05 percent over household consumption percentiles. Over a span of a generation, these differences can create large inequalities in educational access. Lastly, college completion rates have historically been concentrated among the top groups. But more importantly, expansion in college was also captured largely by the top 10 percent despite increases in family incomes over the course of 20 years. The bottom 50 percent had an paltry increase of less than 0.25 percent completion rate per year, meaning it would take over four years for the bottom half to see an increase in college completion by 1 percent. In contrast, the richest households grew by about 2.5 percent chance per year, meaning their annual gains would take the bottom half about 10 years to match up in pace.

2.2.1. Hidden pink tax on educational access. The results are damning. However, our initial time series suggested systematic differences between men and women and between caste groups for any given birth cohort. Therefore, we expect our series to demonstrate a gap between groups at any given time. However, it is not obvious ex ante if these gaps are driven by the bottom income households due budget constraints or if they persist even among the rich, which may highlight systematic differences in occupations and returns to schooling.

We first analyze gender inequality. (see Figure 2.2.3) Women in rural areas face far different constraints in comparison to women in urban areas due to differences in gender roles and occupations. The former are mostly engaged in farm labor and the value of education is evaluated against the opportunity cost. Urban women's education tends to be prioritized by marriage markets and skilled job opportunities. While the opportunity cost argument holds credence for rural men as well, they are also seen as prospective migrants to urban areas to remit earnings. Urban men are educated with upward economic and social mobility as a priority by moving away from unskilled to skilled occupations. Since the rural and urban gender roles are not comparable, we separate our analysis for the two. Having known the pattern of growth in education by household consumption, we only present the distributions for 1983 and 2004. For clarity and comparability, we use the nationally estimated deciles for household consumption, but split the top decile into top 5 and the next 5. Consequently, any gap between the two genders is unexplained by region and income. Figure 6 summarizes our findings for the two years for individuals between 25 and 30 years of age.

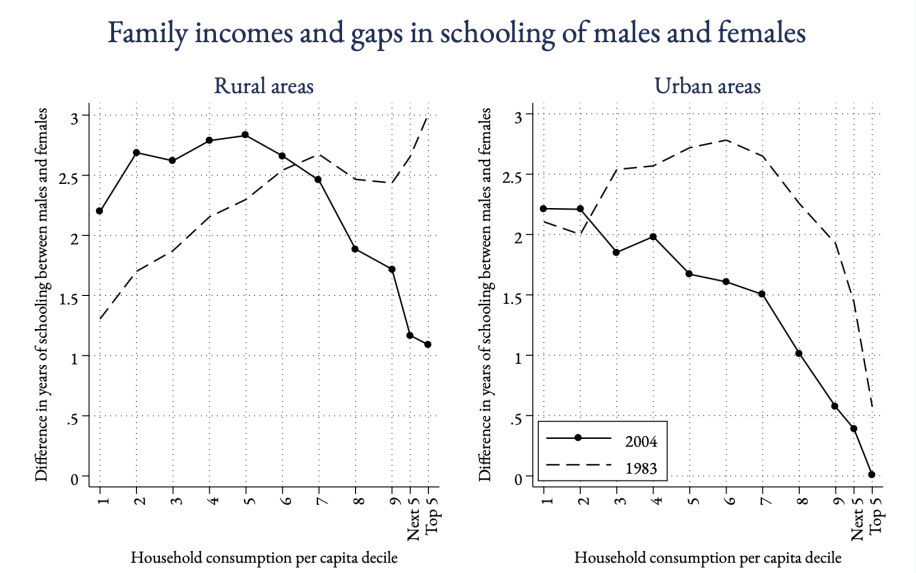
We confirm systematic differences between men and women at every income and education level. The distribution for secondary school completion is analogous to overall schooling, although college completion shows far higher dispersion. As recently as 2004, women lagged behind men by nearly 50 household income percentiles to have an average attainment of primary level schooling and by over 2 years for any given level of income within the bottom



Source: Author's calculations using National Sample Surveys.
Note: Samples are individuals aged between 25 and 30 years. Household consumption per capita percentiles are estimated for full population.

FIGURE 2.2.3. Gender differences in education

80 percent households. While rural women did not even reach primary attainment including among the top groups in 1983, only the top 5 percent of urban women had completed elementary schooling.



Source: Author's calculations using National Sample Surveys.
Note: Samples are individuals aged between 25 and 30 years. Household consumption per capita percentiles are estimated for full population.

FIGURE 2.2.4. Bias in education towards males

However, differences between men and women vary by rural and urban areas. The richest households historically had very little gap between the two the three measures of schooling in urban areas, whereas rural areas have transformed from a society of systematic social differences to one that is also driven by incomes at the very top. Figure 2.2.4 illustrates the difference between rural and urban areas in the gap and the changing relationship over time. The closing gaps between men and women among the richest households in rural areas may be driven by changing occupational structure and marriage markets, but the gaps for the bottom 80 percent may still be dominated by the relative cost argument. The difference in access for identical income levels for each sex between the two regions may be attributed to greater access to schools and colleges in the latter and differences in returns to education. Despite differences up to secondary schooling, college completion is relatively gender equal in urban areas for the recent cohorts—this was not the case in the 1983. This suggests that conditional on completing secondary school, gender differences in access have shrunk for urban women. The rural story, however, demonstrates that the differences in high school have continued to carry over to college.

In all, women are not at par with men in access having accounted for incomes and regional differences. What were once gender based gaps throughout the income distribution

have been complemented with income dynamics as well. Our results suggest that household budget constraints and returns to schooling may play an important role in determining whether women have access than purely differential gender roles. Limited household resources compile parents to invest where returns to education are perceived to be the highest; in a market where men are expected to play a key role in upward mobility of households, these investments occur in men. These differences are what may be considered an analog of the pink tax in educational access: just being a woman rids a person of the preference to have the opportunity to educate themselves.

To eliminate gender gaps, our evidence suggests that at a minimum, when households cannot afford to invest in their children's education due to direct costs, governments need to substitute for the household's contribution; and where opportunity cost plays a role, governments need to support families through financial aid for lost earnings. Although current enrollment data suggest that gender parity has balanced since the recent abolition of child labor and universalization of schooling, it may not be sufficient since the dynamics for overall schooling will likely be replicated for college education. Rural areas demonstrate the largest gaps and constitute 60 percent of the population; therefore, the gender gaps there account for the largest share of lost opportunity. Since gender specific preferences are likely at play as well in contributing to differential access, enrollment drives and schooling supply will benefit from being complemented by girl education specific campaigns in both rural and urban areas. As tautological it may sound, educating women may be the best way to reduce gender parity since a daughter's education is rarely less than her mother's education.[7]

2.2.2. The continuing relevance of caste. A volume of contemporary debate ponders whether caste is still relevant today with anecdotes from either sides. The routine argument for affirmative action is that there continues to remain a *savarna* hegemony due to which SCs and STs do not have opportunity. The argument from the other side suggests that it is merit that separates opportunity and not systemic oppression. Numerous quantitative studies that we reviewed suggest systematic average differences of access between the "reserved" (SCs, STs and OBCs) and "unreserved" (the rest)—in line with the time series of differences in educational attainment by caste—which underpins the value of affirmative action policies to this day. However, there is little evidence on the distributional access of the two groups, let alone its evolution over time. In particular, an important question that remains to ask is whether it is true that incomes should form the basis of affirmative actions on education or whether it should continue to remain heavily caste based. This section examines this question.

Using methods identical to above, we estimate the distribution of education by family incomes using survey data for various caste groups listed in the NSS for individuals of ages

[7]See Figure A.0.2 for educational transmission from mothers that we estimate from National Sample Survey data linking children with their mothers in the household.

25 to 30 for years 1983 and 2004.[8] However, using caste groups to assess differences in education deserves some justification since aggregate differences in caste may simply capture differences by sex and region due to the heterogenous distribution of their population in these areas. First, caste as a category explicitly concerns inequality of opportunity since *Dalits* and the indigenous tribes historically were not a part of the schooling system by virtue of their traditional occupations. Further, unlike geographical mobility, caste is assigned by birth and passes over from family. Second, due to historical structural differences in access, policies on affirmative action are based on caste, regardless of sex and administrative regions. Therefore, the degree of systematic differences in access by caste would highlight its relevance today in comparison to income, which has been a contentious issue vis-a-vis affirmative action policies. That said, SCs and STs share more similarity than OBCs do among the "reserved" communities; OBCs were neither a category before the 1990s and do not share the same historical bases of discrimination that SCs and STs do. For brevity, hence, we aggregate SCs and STs. The distributions of access versus household income are summarized in Figure 2.2.5a and the chance of having a college degree conditional on completing secondary schooling is summarized in Figure 2.2.5b.

While the routine interpretation of convexity in all three types of educational measurements hold, there are a few observations that deserve discussion. First, transition from secondary completion to getting a college degree follows the trajectory of Figure 2.2.1 for all castes. The bottom 80 percent transition less than 20 percent of the time regardless of caste as of 2004 while the increase among the top groups in 1983 was flatter. This highlights potential threshold effects of absolute levels of income in enabling access. The gap between FCs and SC/STs of over 5 percent for each decile in each period may highlight caste specific occupational segregation or systematic barriers to entry to the SC/STs not captured by household income / consumption.

Second, by and large, differences within castes supersede differences between castes. The range in years of schooling between the top and bottom income groups regardless of caste exceeds 10 years as of 2004, the educational equivalent of which is one full generation; on the other hand, the difference between castes falls only within the range of 1 and 2 years for every level of household income. (see Figure 2.2.6a) This trend is in sharp contrast to 1983, where caste gaps in schooling (regardless of how we measure it) increased gradually with income levels. Even though class based inequality of access exceeded caste based inequality in 1983, the difference between the two was marginal: caste played a significant role in differential access in the past, and the systematic differences continue to hold to this day.

Third, while the differences in overall and secondary schooling between castes have reduced among the top groups, an outlier has been college completion where the differences between castes not only increase exponentially over income, but have also increased over time for every income level. College access has been a key ingredient in affirmative action

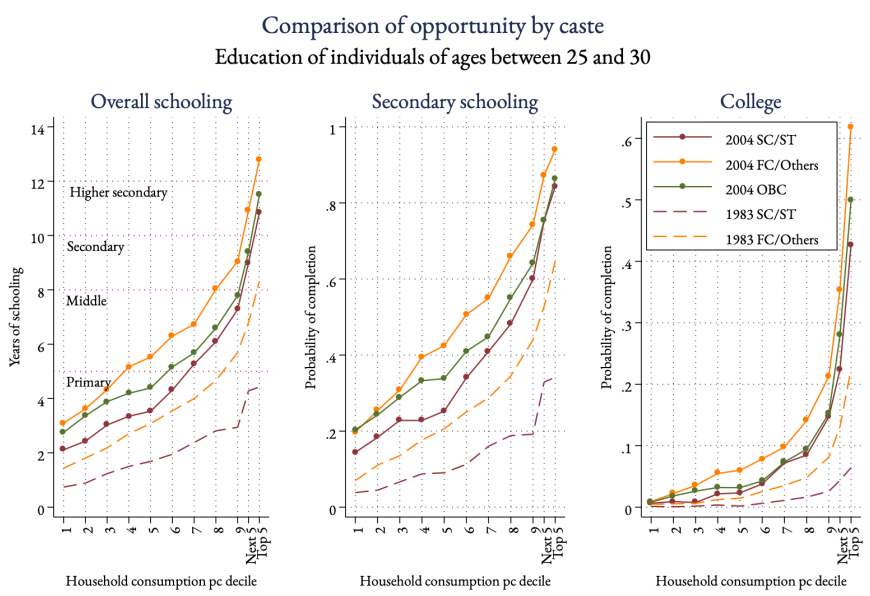
[8]See description of data section for more details.

policies—our evidence suggests that despite policies to increase aggregate representation, gains have been unequally distributed with FCs gaining substantially from an aggregate increase in college supply.

Fourth, bottom and top groups from each caste have benefitted differently over time. (see Figure 2.2.6b) Among the bottom 80 percent, FCs have gained more years of schooling and secondary completion rates in the twenty years since 1983 despite starting from a higher point than SC/STs, whereas the latter have gained more among the top 20 percent. There is no natural economic explanation for why gains should be higher among FCs despite similar household incomes as SC/STs and a higher starting point to begin with. Although occupational segregation and increased preference for occupational segregation may explain the differences in need for skills to compensate for differences in returns to schooling, differences in barriers to accessing schooling may also explain the same. The precise mechanism deserves scrutiny. Since SC/STs gained much more overall schooling than FCs among the top 20 percent, it reflects increased preference or reduced barriers to access schooling due to changing occupational structures at the top with increased competition for common resources in the labor market. Despite an increased access in overall and secondary schooling, college gains by FCs once again dominates those of SC/STs at every income level. There is no doubt, therefore, that college is one domain where SC/STs face systematic differences of opportunity. Since high skill training leads to different job opportunities, our evidence may explain differences in shares of FCs and non FCs in high skilled occupations and, therefore, differences in caste based shares among top incomes controlling for all other factors.

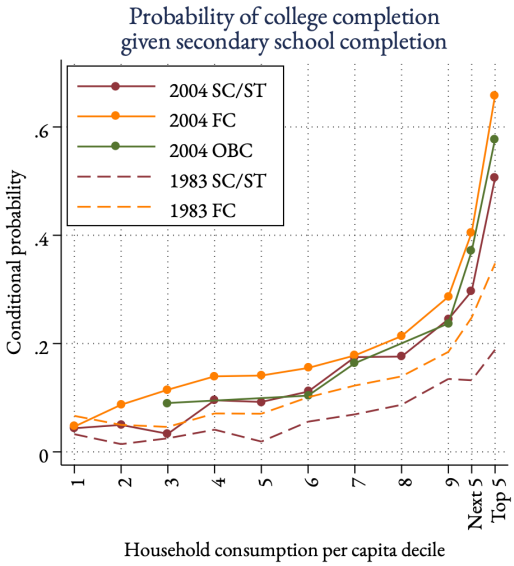
Lastly, OBCs form a peculiar caste category since they share a near identical access with FCs at the bottom of the income distribution but align with SC / STs at the top end. Seen in conjunction with the evidence so far, caste based policies on affirmative action may still be relevant to this day.

The cohorts analyzed range from 1950s to 1980s and form the bulk of adult population today. In light of the fact that differences in returns to skills are substantial between college graduates and high school dropouts, our results provide an accounting for why income inequality in India grew over the recent period. The richest individuals have had a non-marginal privilege in accessing higher levels of education; further, persistent differences in high skill education between castes over time highlights that the cleavages between the FCs and non-FCs have been increasing at the top end. At the bottom too, we see that FCs have benefitted the most over time suggesting barriers to non-FCs in accessing schooling despite having similar incomes. These differences in gains despite low incomes may explain why FCs tend to believe in the narrative of merit, and why they may believe that the richest groups among non-FCs may be unfairly advantaging since the latter's gains have been non-marginally higher over time. However, aggregate evidence suggests that these narratives may



Source: Author's calculations using National Sample Surveys.
Note: Samples are individuals aged between 25 and 30 years. Household consumption per capita percentiles are estimated for full population.

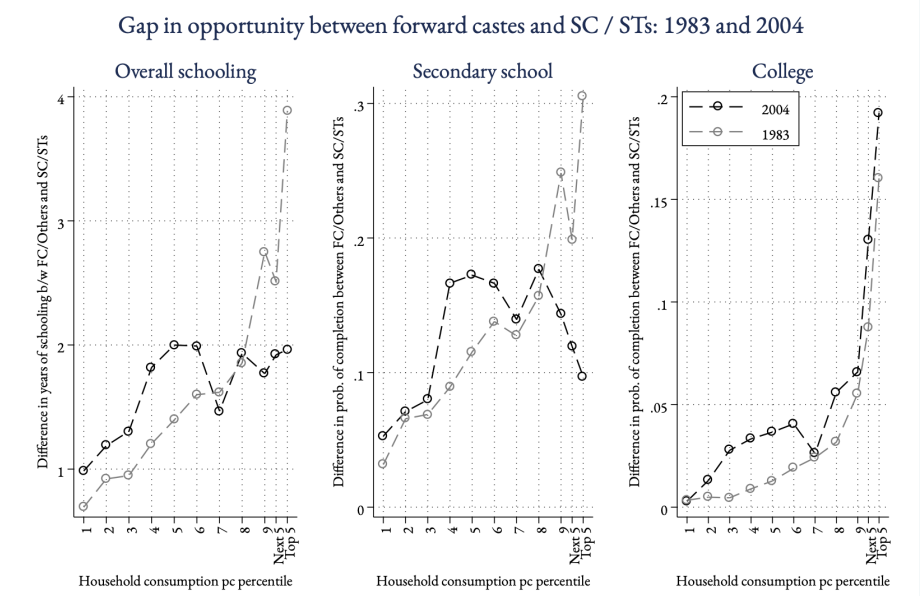
(A) Overall schooling



Source: Author's calculations based on National Sample Surveys.
Note: Each point represents the ratio of probability of college completion and probability of secondary school completion.

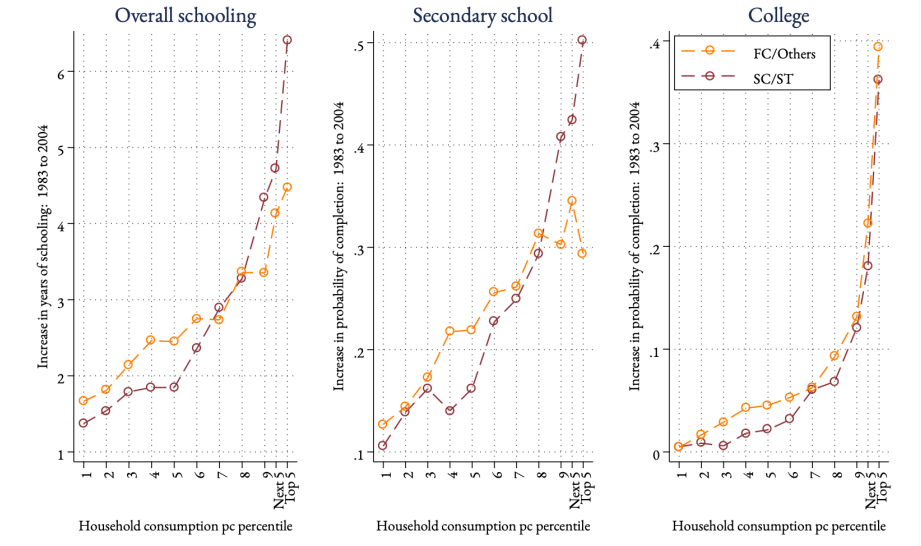
(B) College conditional on completing secondary school

FIGURE 2.2.5. Caste decompositions in school access



Source: Author's calculations based on National Sample Surveys.
Note: Samples are individuals aged between 25 and 30 years. Household consumption per capita percentiles are estimated for full population.

(A) Differences between caste



Source: Author's calculations based on National Sample Surveys.
Note: Samples are individuals aged between 25 and 30 years. Household consumption per capita percentiles are estimated for full population.

(B) Gains for each caste

FIGURE 2.2.6. Caste: within gains and between differences

not hold truism since despite overall closing gaps, FCs remain advantaged at every level. College access in particular sees that the "unreserved" communities have gained far less than the "reserved" ones, and affirmative actions policies hold validity.

Nevertheless, there remains merit in the argument that affirmative action may benefit from incorporating income based criteria since differences within castes (of nearly 10 years between top and bottom groups) exceed differences between castes (of less than 2 years at every income level). Although FCs may have privileged access to schooling among the bottom groups, it may be true that they prioritize it at a high opportunity cost. Therefore, the argument of merit cannot be overlooked entirely. Since it is the SC/STs who have lower levels of schooling among the bottom groups, affirmative action policies combining caste and income may benefit them the most. Although "creamy layer" clauses controls positive discrimination among the richer sections of "reserved" communities, the potentially high opportunity costs borne by FCs at the bottom groups to access education could be alleviated by incorporating income based policies.

In total, we complement existing literature on inequality of opportunity by producing evidence on the recent 30 years in India. We find that educational access is by and large determined by household's access to resources regardless of gender and caste. The lowest income groups as recently as 2011 have as little as 3.5 years of schooling, about 10 percent chance of completing secondary school and about 1 percent chance of having a college degree. The richest households on the other hand complete high school on an average and have a nearly 50 percent chance of having a college degree. Moreover, since 1980s, the gains in schooling and college have been higher the richer the household.

While there persisted systematic differences between the two sexes regardless of relative income about half a century ago, recent data suggests that households virtually do not discriminate between the two sexes among the richest households. Since labor markets are biased in favor working males, gender differences in access may reflect rational choices of resource crunched households to educate males over females. Supply of schools for girls may also drive female enrollment. However, our evidence does discount the possibility of discrimination against females. In either of the three mechanisms driving female education, our results underline the need for reducing opportunity and direct costs of schooling for poorer households in enabling greater access to women. Despite improving gender parity since RTE, government needs to continue expanding schooling for women by increasing school access and subsidizing households.

Further, we find systematic differences between castes at every income level, and particularly pronounced differences in getting a college degree. While this lends credence to the story of systematic caste based discrimination may not be false, we also find some basis for the other side of the debate on the affirmative action since within caste differences exceed between caste differences in schooling, and FCs gained more schooling than non-FCs despite

equal incomes for the poorest households. While indeed the cause for systematic differences between caste may not reflect taste based discrimination nor may the increased schooling among FCs reflect caste-based differences in preferences for schooling or merit of the FCs, what remains certain is that the role household incomes play in accessing simply cannot be negated. Parental income is an integral part of inequality in accruing human capital.

However, communities in India are shaped more by caste and religion than by class due to similarities and widespread prevalence of cultural-religious practices common to castes and religions. Therefore, although there are substantial differences in educational access by class, social and political preferences may not be shaped by them if similar class individuals from one caste identify themselves differently from those from a different one. We know that there have been nontrivial historical cleavages in political preferences between minorities and non-minorities by both caste and religion in India. (see Banerjee, Gethin, and Piketty, 2019) However, the origins of these differences are not well identified. We believe that these differences in preferences may be rooted in how differently the two communities view opportunity. Views on opportunity are formed by the network a child grows up in since the access to opportunity they possess is likely to be perceived as the norm. If networks of caste-religious communities are cleaved, as are say their opportunity, then it is natural for different communities to form different preferences. If the cleavages between communities have historical origins and systematic differences between communities in opportunity persist over time, then it is natural for their views to diverge. Our aggregated category of "FCs and others" reflect too broad a category to draw conclusive caste based interpretations on divergence of preferences since it encompasses various castes and religions. Further, our analysis so far has only concerned the recent period where income dynamics have begun complementing broad caste dynamics. In the next section, we examine the precise historical differences in opportunity between minorities and non-minorities to address some of these limitations and explore the potential sources of divergence of preferences.

2.3. Brahmins and Muslims: 100 years of divergence

> It is a pity that Caste even today has its defenders. The defense are many. It is defended on the ground that the Caste System is but another name for division of labour; and if division of labour is a necessary feature of every civilized society, then it is argued that there is nothing wrong in the Caste System. Now the first thing that is to be urged against this view is that the Caste System is not merely a division of labour. It is also a division of laborers.
>
> — Ambedkar (1936)

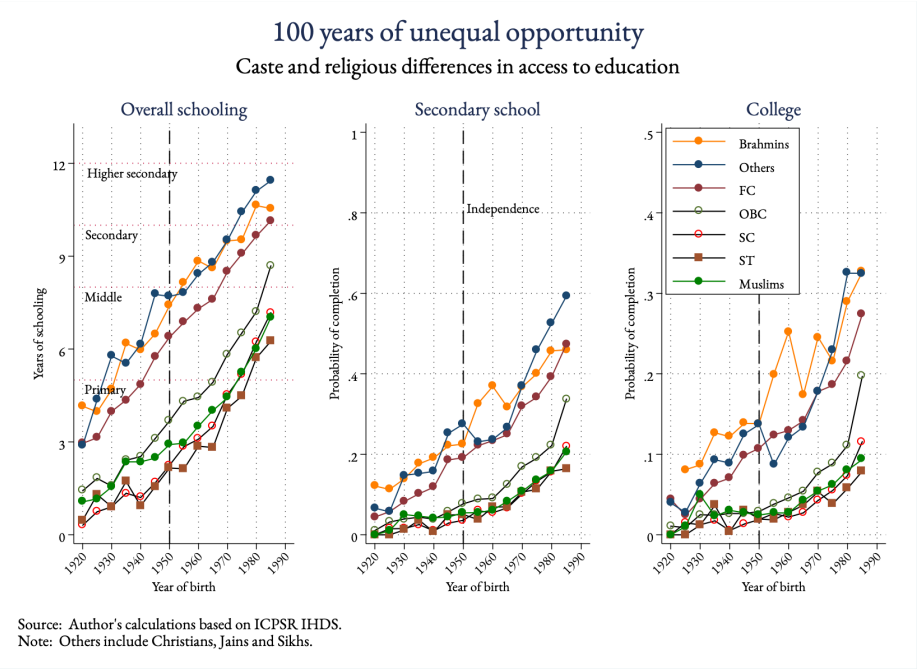
A sensitive issue at the heart of unequal opportunity in India has been the historical divide between castes and religions in India—in particular, Brahmins, who had access to education,

and what today are collectively referred to as the scheduled castes (SCs) and scheduled tribes (STs), who did not. During the colonial rule, *jatis* played a key role in separating traditional occupations, and consequently the skills required in carrying them out. Castes and religions as a crystallized identity emerged slowly as *jatis* began to be aggregated by their occupations during the colonial rule for administrative purposes and population statistics. Brahmins as a collective tended to be priests and scholars, and thus were more literate than the average population. Similarly other forward castes, such as Vaishyas, Kshatriyas and Kayasthas who constituted tradespeople, accountants and administration required numeracy and literacy. Those that carried out vocational trades, inclined towards specialized education in the crafts (such as Muslims), and those oriented towards businesses and high skilled occupations partook in higher education (such as Parsis, Christians and Jains). The remainder occupational groups and labor classes had few opportunities to access education. Figure 2.3.1a demonstrates the basic differences between caste and religious groups over the twentieth century.[9]

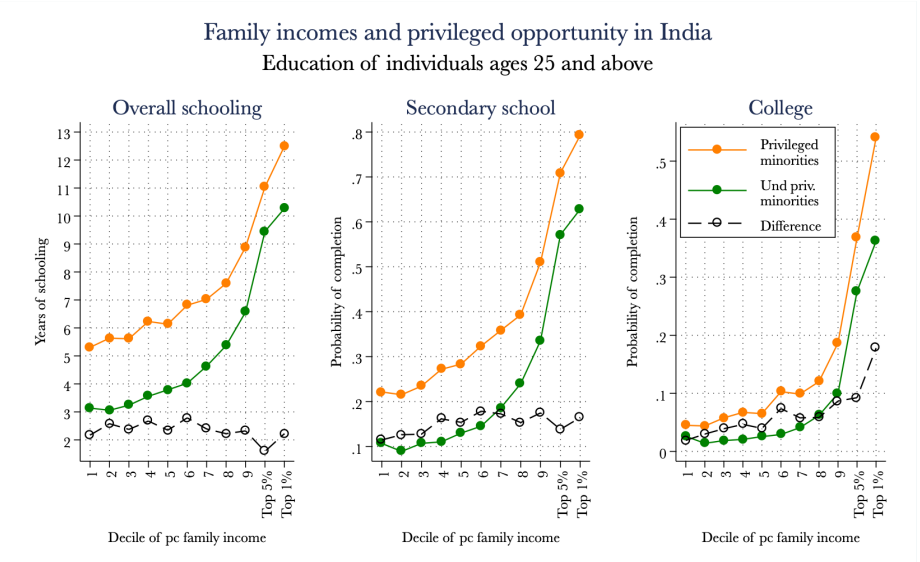
What becomes evident is the cleavage between the privileged minorities (Brahmins, other forward castes, Paris, Christians and Jains) and the underprivileged minorities (Muslims, SCs, STs and OBCs). Privileged minorities (henceforth "PMs") had a consistent advantage of nearly three to four years of average schooling over the underprivileged minorities (henceforth "UMs"). Until as recently as 1980s, while PMs completed secondary school on an average, UPs did not complete elementary schooling. This inequality mechanically translates to an underrepresentation in higher education where we notice a nearly 20 percent gap in the chance of having a college degree between the two.

2.3.1. Incomes and privilege. Indeed while these differences persist between groups, they may be far overshadowed by differences within groups. As before, we utilize the per capita household income as the potential source of variability in educational access within groups and analyze the differences between groups by aggregating the privileged and underprivileged minorities respectively. Figure 2.3.1b demonstrates these differences for individuals of all ages above 25 years by deciles of per capita household income using the ICPSR Indian Human Development Survey data from 2011-12. The cleavages over time are replicated not only on aggregate, but also by the income of the household. There is a systematic gap of over 2 years of schooling on an average between PMs and UPMs for every level of family income at the national level, and an increasing gap for secondary and college education the richer the household—including for the top 5 and the top 1 percent. Completion rates for privileged groups is over double that of the underprivileged groups for the poorest half of India; even

[9]We estimate historical values by back-tracing the closest quinquennial year of birth from survey data. Indeed it is possible that there is a selection bias among the surviving individuals, but we suppose that this selection is unlikely to vary substantially on an average between groups to the extent that the reported values misrepresent between group differences. The type of selection among older survivors is likely to favor those with a higher education given that they may come from a background with higher health provision, and therefore, the average values may, if anything, represent optimistic historical estimates for educational attainment.



(A) Evolution of schooling by birth cohorts: 1880 to 1980



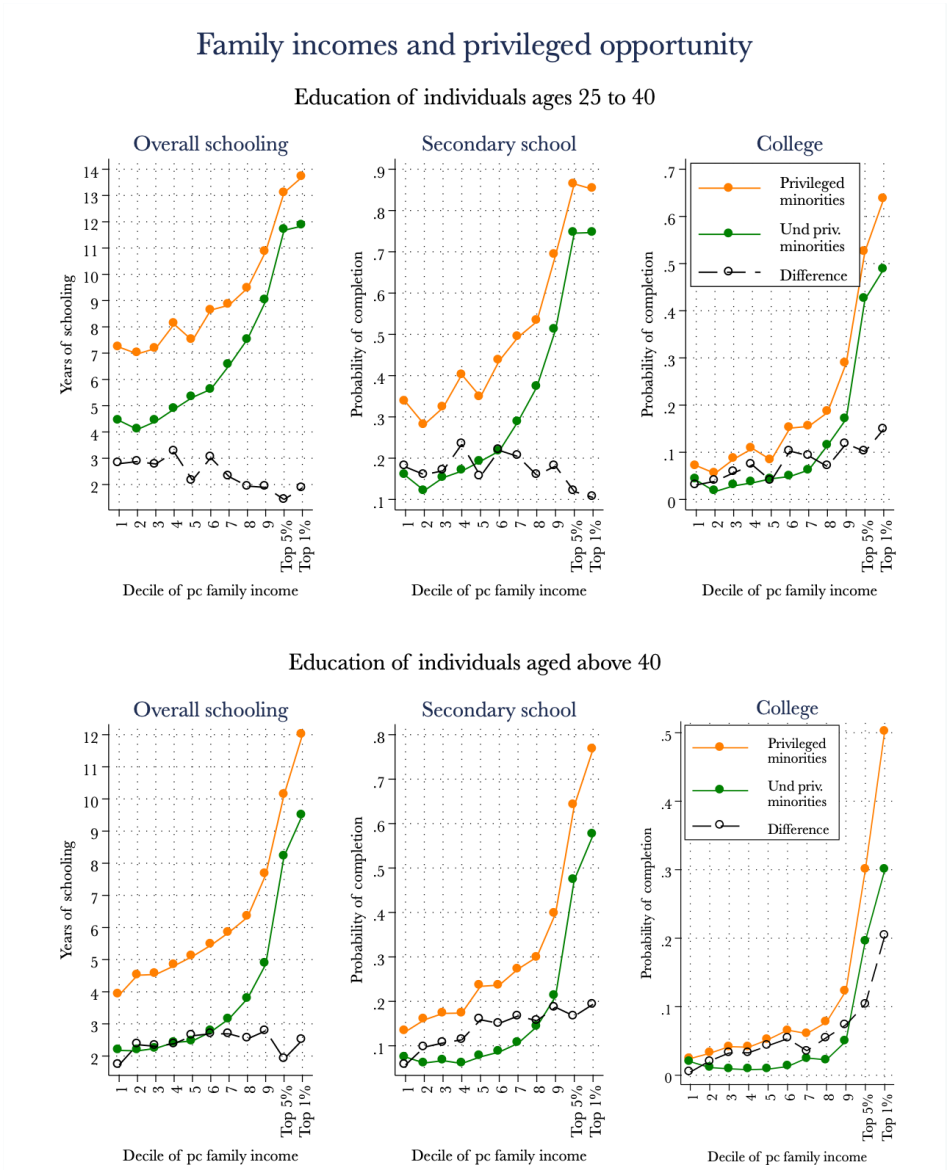
(B) Differences in schooling by family income

FIGURE 2.3.1. Unequal opportunity in India by caste and religion

among the richest 1 percent, completion rates for PMs exceed by 25 percent for secondary school and by 30 percent for higher education.

Individuals of age 25 and above form a majority share of the electorate, and these alarming differences may play a key role in forming social and political preferences based on their experience of access. Recall that the age group corresponds to all surviving as of 2011-12 and born in or before 1985. Those born during 1970s and 80s witnessed the expansion of schools and colleges in the 1980s and 1990s ushered by liberalization; they were also the the last generation before campaigns for universalization affected enrollment. Therefore, the differences in educational access of the younger and older generations may tell us more about the changing trends in difference of access between the privileged and underprivileged groups. Figure 2.3.2 segregates educational attainment for those who completed their studies below the age of 40 and those above the age of 40. First, the difference between the two groups remains high regardless of age differences. The older generations show a constant gap of about 2.5 years in average schooling and over double the college completion rates for the bottom 90 percent population for the privileged. Even among the richest 5 and 1 percent, college completion rates exceed 1.5 times for the PMs in comparison to the UPMs. These differences are depressed among the younger cohorts, particularly the richest households, reflecting an increased access in the 1990s and 2000s—but nevertheless, the gaps remain systematic and significant. Second, the biggest gains in education were accrued the poor among PMs. The magnitude of difference in average schooling between the two groups has increased by about half a year for the bottom deciles, despite the decreasing trend with increase in household income. Third, a more optimistic analysis would suggest that despite lower access to education, the privileged and underprivileged minorities show a similar level of per person household income. This raises an important question about the differential returns to education for the two groups, systematic occupational differences and differences in sources of incomes. Given the diminishing gaps in attainment and preference of the rich UPMs to educate their children, we are left wondering whether education has been playing a changing role in income formation to retain similar levels of income, or whether it reflects social preferences on education as social and cultural capital. Fourth, despite a distinctive gap between the groups, the overall increase in levels and decrease in gaps among the richer households may play an increasing role in influencing political and social preferences by class over generations. However, while this may be true on an aggregate for the country, preferences are likely to be formed at the local level.

To examine the variation in differences between PMs and UPMs, we split our analysis between those states that prioritize social spending and those that do not. In the spirit of Banerjee et al. (2019), we segregate nine major states covering various parts of India by their long run political inclinations and social spending into left, center and right. Despite the fact that the three political labels borrowed largely from the Western political context



Source: Author's calculations based on ICPSR IHDS 2011.
Note: Secondary school completion probability represents the share of percentile population who have completed at least 10 years of schooling. Individuals may or may not have completed high-school which requires 12 years of formal schooling. College completion probability is estimated as the share of population with at least a Bachelors degree or 15 years of formal education. Privileged minorities consists of forward castes, Brahmins, Christian, Jains and Sikhs. Underpriviledged minorities consists of Muslims, OBCs, SCs and STs.

FIGURE 2.3.2. Differences in privilege by age

do not fit perfectly in the Indian context, they do retain broad similarities in ideology and provide interesting insights on differences between states. We classify Kerala, Tamil Nadu and West Bengal into left due to historical rule of socialistic parties; Karnataka, Maharashtra

and Uttar Pradesh broadly into center due to fluctuations in parity of party in power; and Gujarat, Madhya Pradesh and Rajasthan into right due to major influence of center right or right leaning parties.[10] The income deciles are estimated at the national level to eliminate effects of state specific levels of income which may influence average schooling.[11] Moreover, national income percentiles also help in interstate comparability. Figure 2.3.3 summarizes the difference in years of schooling by income levels between states by political ideology. An analog for college completion by state is included in Figure A.0.3 in the Appendix. Kerala demonstrates a remarkable contrast with right leaning states with virtually no gap between the PMs and UPMs. States with a history of political movements supporting Dalits and backward castes, such as Maharashtra and Karnataka similarly show relatively thin gaps. States with presence of historical social elites, such as West Bengal and Tamil Nadu, demonstrate significant gaps among the top 5 and the top 1 percent. The two states with a large presence of a minorities, namely Uttar Pradesh and West Bengal, demonstrate significant gap of about 3 years of schooling for the population and about 5 years for the top 1 percent income households. These may represent substantial inter-caste occupational differences, given a closing gap for the top 5 percent which may represent a managerial class requiring higher levels of education. The closure of educational gap among the top 1 percent households in Kerala, the only state where gap does not increase at this income level may reflect the lack of caste based occupational differences.

The results so far demonstrate two key features of social bases of unequal opportunity: first, that there are significant differences by caste and religious identity, and second, that these differences differ by state for those who have completed their schooling. However, what does evidence have to tell us about those currently enrolled? There are reasons to believe that the drive to increase elementary school enrollment by campaigns to universalize education in late 1990s through 2000s played a role in reducing inequality as did greater expansion of colleges and universities. As a first order exploration, we analyze the share of enrollees by caste-religious groups of India over different levels of schooling and contrast of these values separated by almost a hundred years. We utilize PEII statistical tables for 1917 and ICPSR IHDS for 2011 for the calculation and present results in Figure 2.3.4. A remarkable similarity is noticed between the two: the share of Muslims and other underprivileged minorities fall as levels of schooling increases whereas the enrollment share of Brahmins, forward castes and other privileged minorities (including Christians) increases. While the evidence for 2011 data is directly implied by previous figures, the similarity with 1917 data indicates that these

[10]Uttar Pradesh's status is precariously center given its inability to be classified as left (for having prominent presence of historical Hindutva) and right (due to minority based politics and social programs). A similar but opposite justification may be given for Maharashtra, where on the one hand, despite emergence of Hindutva in the state, the presence of its strong flavor in state level politics has been largely absent; on the other hand, Maharashtra has also been the site of emergence of Dalit and subaltern movements.

[11]For instance, a poor state's top percentiles may have low levels of schooling whereas a rich state's bottom percentiles may have higher levels of schooling.

Source: Author's calculations based on ICPSR IHDS 2011.
Note: Privileged minorities consists of forward castes, Brahmins, Christian, Jains and Sikhs. Underpriviledged minorities consists of Muslims, OBCs, SCs and STs. See text for explanation on separation of states by ideology.

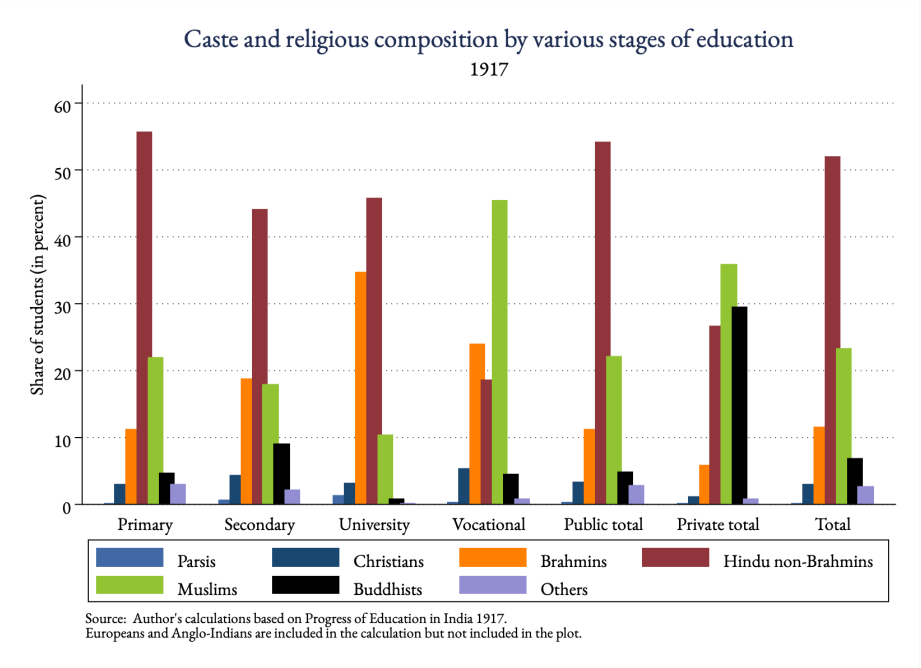
FIGURE 2.3.3. Differences in privilege by state and ideology

differences have persisted since the colonial period. What is further remarkable is that these similarities extend to non-Brahmin Hindus who constitute a stable share of over 40 percent in colonial India over all levels of schooling and OBCs in recent period who constitute about

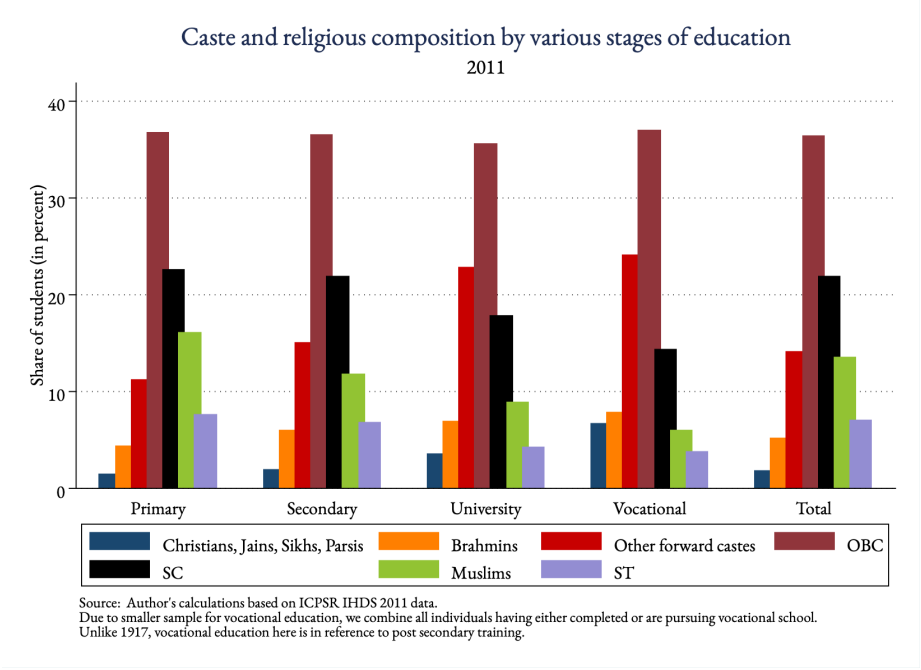
35 percent in all levels of schooling. These patterns suggest that there may be key forces at play during primary school that inhibit progression of the underprivileged minorities and promote progression of privileged minorities.

Furthermore, Muslims occupy a majority share of enrollment in non-college vocational or training schools in colonial India in contrast to the recent period. The falling share of Muslims in trade-schools may reflect a number of mechanisms including migration of skilled Muslim labor to Pakistan and Bangladesh, lack of formalization of vocational schools at the same rate as general schooling, progressive lack of investments in quality vocational education, or simply increased social value of general schooling and university due to elite followership. Regardless of the mechanism, the avenue for Muslims and other UPMs to represent themselves in education seems to have diminished and the progression of schooling system may have come with no harm for the privileged minorities over time.

It is certainly not evident if PMs have not been harmed, since the absolute share of Brahmins, for instance, has indeed decreased at every level of education. To test whether PMs have been harmed, we estimate the overrepresentation for each caste-religious group as the ratio r of the share of all enrollees in a formal educational institutions at or below the age of 25 and the share of population by the same group and index it to 100. A group whose ratio is $r > 100$ is overrepresented in enrollment and those with $r < 100$ are under-presented. (For reference, a group is twice as represented if $r = 200$ and half as represented if $r = 50$.) Since different provinces of India had varying shares of caste-religious populations, overrepresentation is best tested by comparing provinces. Figure 14 describes the variation in overrepresentation r by states for all provinces of British India for 1917 using the PEII and major states or state-groups in 2011 using ICPSR IHDS data and is ordered by representation of Muslims. We do not present extreme values to retain integrity of the plot. In 1917, privileged minorities (Brahmins, Christians and Parsis) are consistently overrepresented, minorities (such as Buddhists) are consistently underrepresented, Non-Brahmin Hindus tend to be slightly underrepresented everywhere and Muslims are underrepresented in some provinces but overrepresented in others. At the country level, although Muslim representation is balanced, non-Brahmin Hindus and others, such as Buddhists, suffer. Burma is the only region where Hindus (including Brahmins) and Buddhists are underrepresented but Muslims are overrepresented. By 2011, the variation and magnitude of overrepresentation has substantially diminished for every state. What's fascinating is the similarity in pattern for Muslims, even though the geographical correlation in representation is low, and the fact that Brahmins are underrepresented in nearly half the regions analyzed, although overrepresented nationally. Moreover, the states with the lowest Brahmin representation are left leaning and the states with highest more or less right leaning. Brahmin overrepresentation at large is weakly, but inversely correlated with Muslim representation. Non-Brahmin forward castes show patterns analogous to colonial period with near proportional representation, and Christians



(A) Enrollment shares in 1917



(B) Enrollment shares in 2011

FIGURE 2.3.4. Replication of caste-religious enrollment patterns

show over-presentation in several states. Moreover, the population proportional representation in schooling do not correlate by province-states over time. This may in part be due to migration of educated Muslims to Pakistan and Bangladesh, although the precise mechanisms deserve closer scrutiny.

These data raise a question: how is it possible that Brahmins are underrepresented in education in so many states when evidence suggests that they in fact consistently have higher levels of schooling? An answer may lie in the choice of denominator in our ratio r. Differing relative fertility rates between communities affect demographic structure of younger and older populations: falling relative fertility implies a relatively aging community and is likely to mechanically inflate denominator for one community in comparison to another and decrease r. Another factor that may influence r is migration among older individuals. Therefore, Brahmin underrepresentation or Muslim overrepresentation may simply be an artifact of changing relative fertility or migration for instance. To discount these factors, we reconstruct Figure 2.3.5 with only the school going age appropriate population (3 to 25) as the denominator. Figure 2.3.6 presents the new values with disaggregations for both religion and caste. The results change dramatically with Muslims being the only major religious group to be consistently underrepresented in schooling, including for India at large, and privileged minorities the only ones overrepresented. There is a consistent 15 percent excess representation by Brahmins and nearly 0 by non-Brahmin Hindus. The latter are virtually the only group with proportional representation in population and school in every province of India. Furthermore, although the share of left oriented states is higher in above average representation of minorities than those below, the distinction is fuzzier than before: Gujarat performs better than Kerala and West Bengal in representing Muslims and has a similar proportion for Brahmins, while centrist states that showed a closing gap for PMs and UPMs such as Maharashtra and Karnataka by income groups show sizable differences. Christians are highly overrepresented Kerala and Chhattisgarh whereas Muslims in these states enroll less. Tamil Nadu and Delhi are the only two regions with negligible variability by religious identity.

However, when religious identities are disaggregated by castes, representation is visibly more dispersed: forward caste non-Brahmin Hindus are overrepresented by almost 10 percent everywhere, OBCs by about 5 percent, SCs underrepresent slightly by about 5 to 10 percent and STs sizably underrepresent *a la* Muslims. Since a number of Christians are STs, separating by caste suggests that non-ST Christians are generally more overrepresented everywhere except Delhi and Tamil Nadu (where STs outperform non-ST Christians). The enrollment patterns as a share of age appropriate population shows patterns identical to anecdotal narratives about identity hierarchies in education. However, what remains far less obvious is the ambivalence in representation between states with presence of social programs or right wing ideology and those that do not. This suggests that despite equitable achievement for individuals who have completed schooling, recent developments may not have benefitted

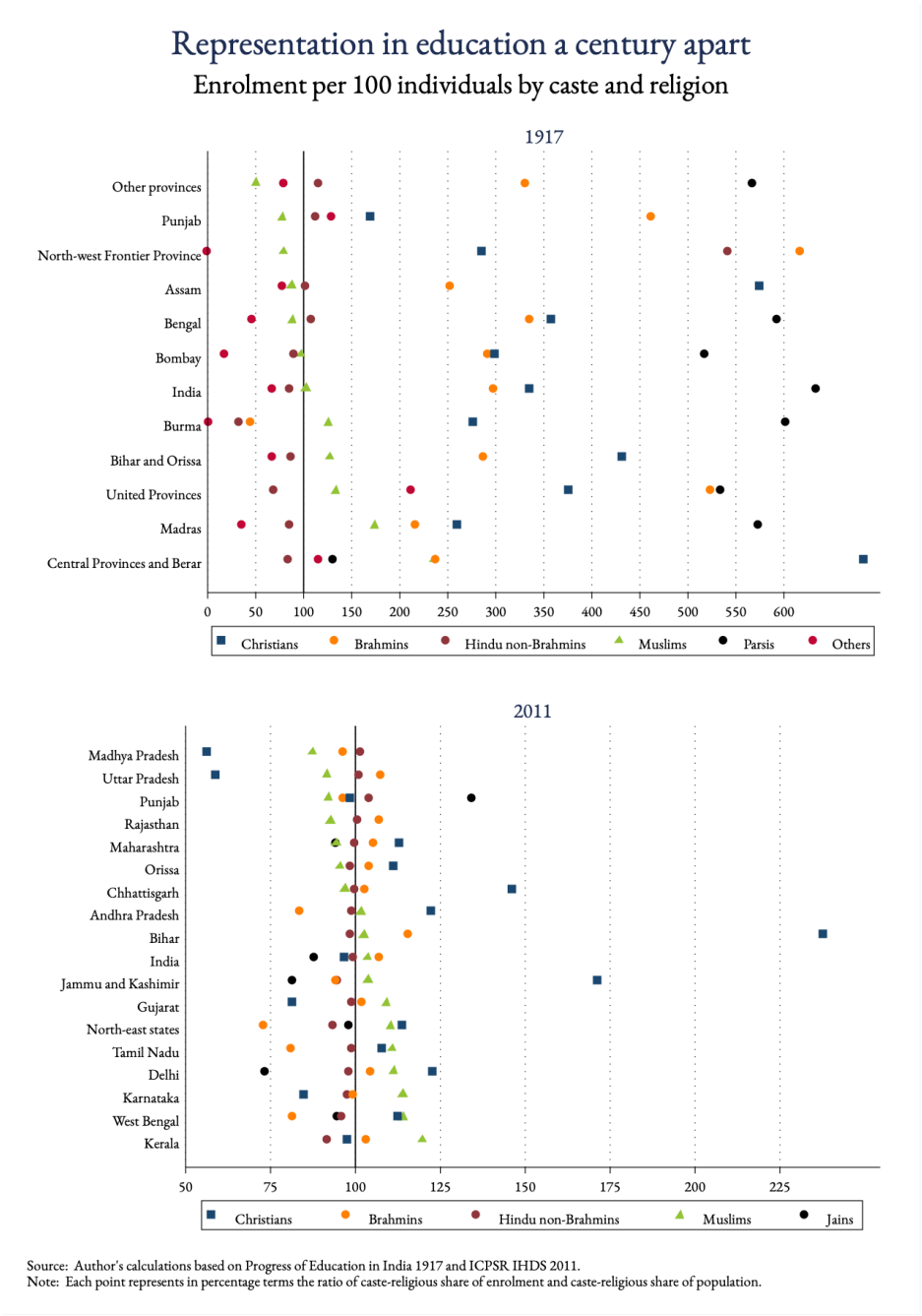


FIGURE 2.3.5. Representation in enrollment with respect to total population

underprivileged minorities the most. These results are robust to alternative denominators such as population of ages 5 or 6 to 25 (not presented for brevity) and therefore are not an artifact of age wise sampling in surveys.

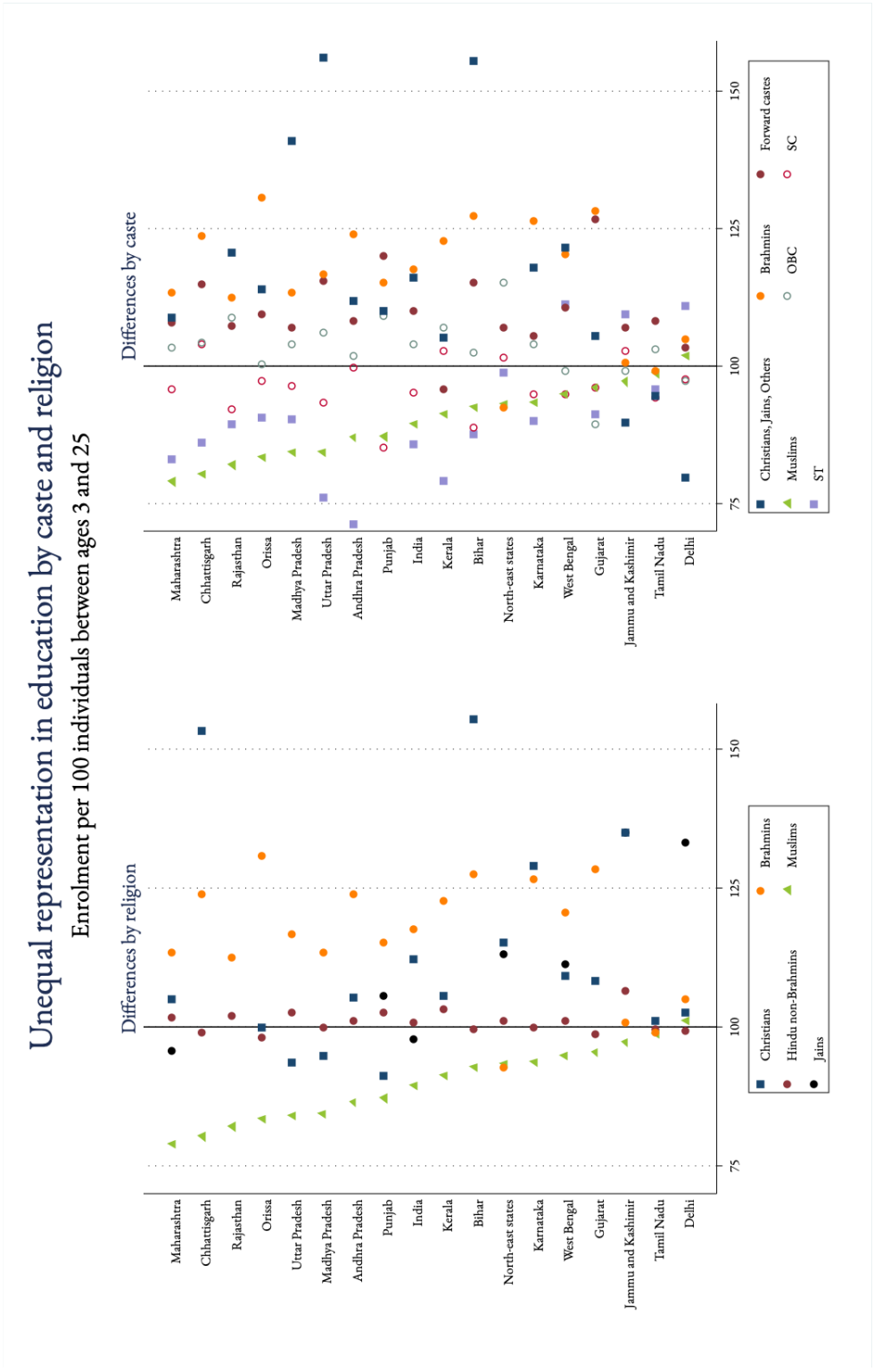


FIGURE 2.3.6. Representation in enrollment by age approrpriate population

The salience of age appropriate ratios in contrast to the population ratios is that representation of a community in education may not reflect true values when interpreting the 1917 plot where we lack age appropriate population by provinces for each religious community. It is well possible that the results are depressed for UPMs once we account for relative differences in fertility between religious groups, even if not fully explain the overrepresentation of Muslims in some provinces. That said, what remains clear is that it is during primary school that divergences are created between caste-religious groups in education. To understand the source of impediment in persistence, we trace primary and secondary stage enrollment by grade level: for every 100 enrollees in grade 1 of a community, how many persist up to completion of high school? Due to lack of longitudinal data in the colonial period, we trace among the only available cross sections by religion and index school enrollees for all grades relative to those at grade 1 in multiple of 100 from PEII 1937 for British India. For comparability, we use the cross section from DISE unit data of 2005 for pre-Right to Education 2009 period and 2016 for post-Right to Education period at the national level. Figure 2.3.7 compares persistence in colonial India, pre-universalization and post-universalization of education for caste-religious groups of India.

For majority of the population, the drops in enrollment follow a pattern of logarithmic to linear to exponential over time. While the reason for drop out post elementary schooling are understandable in post-RTE era, given completion for compulsory schooling, the drop after just one year of schooling in colonial period is not. Consistent with evidence so far, a key takeaway is that Christian and Hindus follow identical trends in colonial period (among which we expect Brahmins to demonstrate higher persistence), and for recent years, forward castes not only show the highest persistence, but also a low drop out rate in elementary school in 2016 of less than 10 percent. This remains in sharp contrast to Muslims in colonial period and STs in recent period, who represent nearly half the share of enrollment in comparison to privileged minorities by the end of primary school despite a law mandating age appropriate enrollment for all children up to 14 years of age. Despite absence of religious data for grade-wise enrollment, evidence so far would suggest that Muslim enrollment would follow somewhere between STs and SCs, meaning over 20 percent difference for every grade level in persistence in comparison to FCs. Parsis are a demonstrable outlier in the colonial period with remarkably high and linear persistence: over 30 percent of all enrollees in grade 1 completed grade 12, and about double that of Europeans and Anglo-Indians, an already elite community in colonial India. In the absence of a strong migration shock during independence, recent data are likely to suggest patterns higher than those for forward castes.

Indeed decomposition of persistence by states is likely to show variation, and the values in Figure 2.3.7 are mechanically weighed more heavily by populous states that may incline towards one ideology or the other. However, the broad story does not change substially given

underprivileged minorities are underrepresented in current enrollment for every level after primary.

Why do underprivileged minorities drop out as schooling progresses whereas privileged minorities stay despite same levels of income? What incentives or constraints bind underprivileged minorities that affect the privileged minorities less in accessing opportunity? What are the underlying forces remain similar from the time India was a colony to one that made education a fundamental human right? While evidence in this section does not answer these questions, the last hundred years of data ascertains that these questions remain relevant even today. The historical cleavages in educational access between Brahmins and Muslims [12]may have moderated, but they continue to remain systematic. The privileged minorities demonstrate a systematic gap of over 2 years of schooling not only on an average over time, but also for every level of income in the recent time. While traditionally left leaning states exhibit smaller gaps between the two communities in comparison to those inclining towards right for identical levels of income, underrepresentation of UPMs in current enrollment data suggests that these variations may be fuzzier, even if true on an average.

The converse interpretation of lower educational access for identical levels of income among older demographic suggests systematic differences between communities in returns to schooling and the role of education in income formation. Preference for schooling revealed by diminishing gaps between the PMs and UPMs among rich households—particularly the top 5 percent—for younger cohorts suggests that formal education may play an increasing role in determining economic opportunity either through skills or through cultural capital due to decreasing relative preference for traditional occupations (or increasing relative preference for identical jobs). The top 1 percent, however, remain less affected; this may be explained by higher share of capital than labor income at the fractile.

2.4. Summarizing access

First, from less than a year of schooling on an average for birth cohorts from 1880s to completing elementary school today, education rapidly expanded over the last 150 years. This expansion was unequal between men and women, urban and rural areas, and between the forward castes and the rest, aligning with the conventional story. A large part of inequality of opportunity is explained by income differences than social differences. The top 1 percent income households exceed the bottom by over 10 years of schooling and over 40 times in college completion. Income thresholds may play a key role in determining college completion given bottom 80 percent have more or less equal chance conditional on completing secondary school. Decompositions of access by caste and religion reveal systematic differences in educational access over about 2 years between underprivileged minorities (UPM)—consisting of

[12]Or more broadly, the privileged minorities that also include forward castes, Christians, Jains, Parsis and Sikhs, and the underprivileged minorities that also include SCs, STs and OBCs.

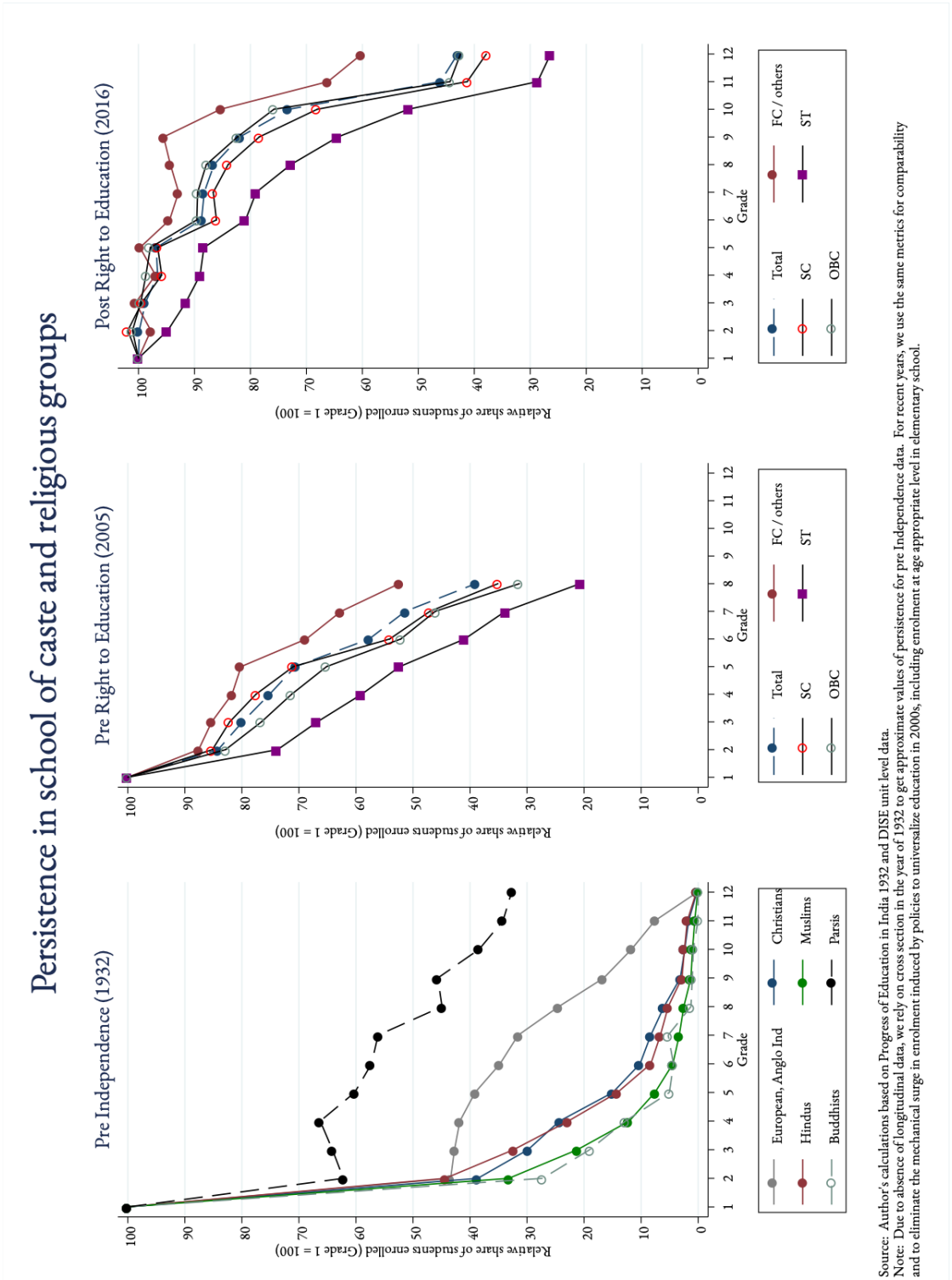


FIGURE 2.3.7. Persistence in school of various caste-religious communities

Muslims, Scheduled Castes, Scheduled Tribes and Other Backward Castes—and privileged minorities (PM)—Brahmins, forward castes, Jains, Parsis and Christians—for every level of family income in any given year, including the top 1 percent. Moreover, the PMs and UPMs replicate patterns of attrition and persistence in school separated by almost a century. Including current enrollment data, the latter continue to remain the only underrepresented group today in every state, whereas the former the only ones overrepresented.

Our evidence for long run and persistent differences in educational access between PMs and UPMs that transcends class may explain why the two may form divergent views of opportunity. While initial differences in schooling may be driven by the different needs for formal schooling towards caste separated occupations, instances where the two groups may compete for common resources (say in the labor or higher education market) is likely to induce the attribution of success to merit for PMs due to higher levels of schooling for the same level of income in comparison to UPMs at large, and discrimination in UPMs for their lack of proportional representation for having lower educational access despite similar income. Since the two sets of communities have followed distinct trajectories in education over time and over income at any given time, and since their patterns of representation in education have remained nearly unchanged for over hundred years, their divergent views of opportunity are likely to translate to divergences in political preferences over ability and discrimination. Since left parties support social spending and target policies to reduce inequality, whereas right parties incline towards the idea of merit in explaining distribution of outcomes, the party compositions of the two are likely to be cleaved by caste and religious identity in the case of India. In particular, we find an argument to suggest why UPMs may vote left and why PMs support right on the basis of systematic differences in opportunity. Closing educational gaps among the richest UPMs and PMs may suggest that the caste-religious homophily may be substituted partly by class homophily among the top 5 and top 1 percent.

Unfortunately, the gaps in schooling are likely to harm the communities engaged in traditional or low skill occupations in coming years. Technological change and the the increasing role of markets and globalization are likely to bias labor markets in favor of high skill jobs and therefore advantage thesis with higher levels of schooling. In our case, this community would comprise of Brahmins, Jains, Christians and other forward castes. Since community networks are likely to be driven in India by caste and religion rather than class, the distinctive advantage of PMs is likely to support them further in mobility in comparison to UPMs through greater availability of information on know-what and know-how. Therefore, there is an urgent need for expansion of schooling among the poorer sections of UPMs to ensure that their dropout rates reduce and transition from secondary to higher education increases. With universal access to elementary school in recent years, focus on improving educational quality for Muslims, SCs and STs is likely to contribute not only to their individual mobility, but also provide a cost effective way to improve economic growth and reduce inequality.

CHAPTER 3

Top shares in educational investments

> To sum up, historical experience suggests that the principal mechanism of convergence is the diffusion of knowledge... Above all, knowledge diffusion depends on a country's ability to mobilize financing as well as institutions that encourage large-scale investment in education and training of the population while guaranteeing a stable legal framework that various economic actors can reliably count on... The educational system must increase its supply of new types of training and its output of new skills at a sufficiently rapid pace.
>
> — Piketty (2014)

Education is a key gateway to economic prosperity, and like any investment, it cannot be accrued without funding it. Naturally, differential spending in education is likely to translate into differential human capital accumulation, both in quantity and quality; and differential accumulation feeds into economic inequality through labor market and institutions. Therefore, it is unsurprising if lack of equitable investments in education impedes mobility before education is even acquired. Worse, when educational supply is outpaced by growth in labor substituting technology, the poorest and least schooled are most likely to be left behind in climbing the economic ladder. (Goldin and Katz, 2009)

There has been a long debate in India about providing equitable access to quality education, and evidence shows that India has fallen short from achieving it by no small margin. The Kothari Commission's recommendation (Kothari, 1970) of investing public funds in education to the tune of 6 percent of GDP has hitherto never been achieved, and the long term lack of equitable financing is reflected in, for instance, over 50 percent of children not achieving grade 2 level proficiency in mathematics by grade 5 as recently as 2010. (ASER, 2010) India continues to have low government spending in education (about 3.5 percent of GDP) despite repeated calls for increased financing (Tilak, 1997a, 2006, 2007, 2010) and faces the challenge of a rising inequality. (Chancel and Piketty, 2019)

Unfortunately, little is known about the evolution of distributional educational supply and spending. In this chapter, we produce the first estimates on concentration of human capital investments in India for two key periods: colonial India from 1870s to 1910s, and Independent India from 1980s to 2010s. The two periods separated by a century play an important role in Indian history: the first marks the rise of formal schooling in India through

European institutions and the second marks the transition to universalization of elementary schooling. The first section of this chapter is devoted to developing a method to estimate cohort wise concentration of educational investments. The second section summarizes the evolution of concentration of investments. We rely on the administrative data archived in the statistical appendices of Progress of Education in India and British Indian censuses for the first period, and the Education rounds of National Sample Surveys for the second. The concentration for the former is estimated in terms of rank of education level, whereas in the recent period, we use rank of family expenditure.

We estimate that a steady 40 percent of all spending on education was concentrated among the top 1 percent of Indians during the colonial period since 1870s, and about 10 percent among the top 1 percent of enrollees. The investments in bottom half of all enrollees nearly equalled that of the top 5 percent. This suggests that a large amount of resources were expended on students who tracked through the full course of schooling and not on expanding mass access. In the recent period since 1980s, we find that inequality in both private and total spending has increased. The share of spending in top 5 and bottom 50 substitute each other. In private spending, the latter decreased while the former increased by 10 percent from 1986 to 2014; in total investments, the latter decreased and the former increased by about 8 percent from 12 to over 20 percent. These numbers indicate a very high level of concentration and are likely to contribute to further rigidification of limited opportunity. The recent equality in concentration of total investments in the top 10 and bottom 50 percent households indicates an equality was last seen in the colonial period in 1870s.

Since states with left leaning ideology tend to spend more than right leaning states on social programs, we decompose our estimates of investment for the recent period by states and ideologies. We find virtually no difference between states in terms of concentration of both public and total educational investments. This suggests differences between states may be guided more by ideological lines on providing a slightly better access to minorities on an average rather than bridging the class divide within states with equitable investments. Moreover, we find that the role of the state in compensating for the inequality created by private investments has increased over time. Therefore, policy proposals calling for increased privatization to expand schooling among the bottom groups may in fact exacerbate inequality if not regulated. We propose that the role of the state is best increased, and the prescient recommendations of Kothari (1970) that have repeatedly made rounds in literature be implemented.

3.1. Concentration by cohort: the longitudinal problem

A challenging problem is to estimate the concentration of educational investments by cohort, for which we do not have direct data: neither for the population or schooling, nor for expenditure. To estimate concentration by cohort, we need two ingredients: first, what fraction of a cohort attended a given level of education, and second, what was the total expense

Educational attainment level	Number of students at level	Composition of life course attainment by level					Expense at level	Expense per student at level
No school	$N_t(u)$	N_t^c	N_t^h	N_t^m	N_t^p	N_t^u	—	—
Primary school	$N_t(p)$	N_t^c	N_t^h	N_t^m	N_t^p		$E_t(p)$	$\overline{E}_t(p) = E_t(p)/N_t(p)$
Middle school	$N_t(m)$	N_t^c	N_t^h	N_t^m			$E_t(m)$	$\overline{E}_t(m) = E_t(m)/N_t(m)$
High school	$N_t(h)$	N_t^c	N_t^h				$E_t(h)$	$\overline{E}_t(h) = E_t(h)/N_t(h)$
College	$N_t(c)$	N_t^c					$E_t(c)$	$\overline{E}_t(c) = E_t(c)/N_t(c)$
Life course educational expenses: (sum of expenses by column)		E_t^c	E_t^h	E_t^m	E_t^p	—		

TABLE 3.1.1. Accounting for educational expenses by time and cohort

incurred for the cohort's education at a given level. We first begin with the assumption on how we would estimate concentration assuming we had this data, and then we describe how to construct the lacking data.

Suppose that N_t^i are the number of individuals born in year t who attain educational level i. Then $N_t = \sum_i N_t^i$ are all individuals belonging to birth cohort t, $n_t^i = N_t^i/N_t$ are the share of individuals with educational attainment i, and $N_t(j) = \sum_{i \geq j} N_t^i$ are the number of students who attended education level j. This also means that N_t^i can be recovered using $N_t^i = N_t(i+1) - N_t(i)$, where $N(i)$ is the enrollment at level i. If p corresponds to primary, m to middle, h to high school and c to college, then, for instance, there were $N_t(p) = \sum_{i=p,m,h,c} N_t^i$ students in primary school, whereas only $N_t(h) = \sum_{i=h,c} N_t^i$ attended high school classes. Indeed, this includes individuals who do not go to school.

Similarly, let E_t^i be the total educational investment for cohort t who attain education level i and let $E_t(i)$ be the educational expense of all students during level i. We know that those who attend college also complete all pre-college studies, those who attend high school also attend middle school, and so on. Therefore, the life-course of educational expense of an individual who attends a level of schooling includes all expenses incurred in their previous studies. If $\overline{E}_t(i) = E_t(i)/N_t(i)$ is the per student expense at level i of birth cohort t, then the life course expense for a student who completes j level of schooling is $\overline{E}_t^j = \sum_{i \leq j} \overline{E}_t(i)$ and the total expense of all individuals attaining i is

$$E_t^i = N_t^i \overline{E}_t^i \tag{3.1.1}$$

Therefore, the total educational investments of a cohort t are $E_t = \sum_i E_t^i$.

The table above helps illustrate the accounting of educational expenses by cohort, where we assume that no expenses are incurred for those with no schooling. Once the number of individuals of a birth cohort attending a level of schooling i is known, it is possible to find the number of individuals attaining i over their life course, and once once the expenditure for a level of education i is known in addition, we can find the total educational expenditure incurred on any population sample. Indeed, by order of education, the top percentiles are

individuals who complete college education and the bottom are those that did not attend school. By ordering all individuals of a cohort by level of attainment, it is possible to find percentile wise educational expenditure. What remains of our interest is the concentration of human capital investments and the contrast between the top and bottom groups. The concentration of investments for top x percent born in cohort t can be estimated as

$$E_{t,p'1-x'p100} = \frac{1}{E_t}\left(w_t^c\overline{E}_t^c + w_t^h\overline{E}_t^h + w_t^m\overline{E}_t^m + w_t^p\overline{E}_t^p\right) \tag{3.1.2}$$

where w_t^i's are weights such that we first count expense of all college students, then high school, then middle school and then primary. The concentration of investments at the bottom can be estimated in the exact opposite manner, where we begin with the population of unschooled individuals. The weights are determined by the population of reference. We estimate concentration for two denominators: for the full population and for the population of enrollees.

Unfortunately, we face at least three technical challenges in estimating this value. First, we simply do not have direct access to the cohort level data on neither education nor expense: at best, we have cross sectional estimates on aggregate level for institutions by year. Due to a number of overlapping generations (5 in primary stage, 3 in middle state and 3 in high stage), and different shares of individuals attaining different levels of schooling with varying expenditures by stage, we need further decomposition of cross sections to trace cohorts. Second, we do not have expenditure data by level of schooling for each year, but rather for educational institutions by type (primary, middle, secondary, etc.)—a secondary school can also have primary and middle students, whereas a primary school has only primary students. Therefore, this data also cannot be used directly and needs decomposition to extract expenditure by stage and not by type. Third, to estimate concentration of investments, we need reliable estimates of not just the enrollees but also those who do not attend any school from the cohort, namely over 85 percent in colonial India. The following sub sections are devoted to resolving these difficulties.

3.1.1. Population estimates for various ages. To estimate the total number individuals who do not attend any schooling, we need the effective number of individuals who were born in year t. Although it is possible to take the simple difference in population between two consecutive years, it may not tell us much given changing age structures. Therefore, for our purpose, we use the decennial age-structure data from the British Indian Censuses, Mukherjee (1976) and McAlpin (1983). To estimate the number of individuals in a given year within the 10 year period, we use the calculated annual growth rate for the age appropriate population to infer population every year in between. If $N(t, a)$ is the population of individuals in an age-band a reported at time t, then for a year t' such that $t \leq t' \leq t + d$, where d is the gap

at which data are observed, the population of age-band a is

$$\begin{aligned} N(t', a) &= N(t, a)\left(1 + g_{t,t+n}\right)^{t'-t} \\ &= N(t, a)\left[\frac{N(t+10, a)}{N(t, a)}\right]^{\frac{t'-t}{d}} \end{aligned} \tag{3.1.3}$$

Now, a cohort whose data is observed at year t' was born in the range $t' - a$. If the size of age band a is more than 1 year, then $t' - a$ and $t' + 1 - a$ overlap for at least one year as do $t' - a$ and $t' - 1 + a$, and therefore depending the initial year of reference, we may get multiple estimates of cohort sizes for each year. Given mortality rates in the first 15 years of life, famines, wars and diseases in the nineteenth century, the differences in estimations can be sizable depending on which age band population is backtracked. As a stable benchmark, we backtrack population of ages 6 to 10 (primary school going age) and for experimental purposes compare it with the backtracked population of ages 11 to 14 (roughly middle school going age). To do so, we first estimate the population of age band wise population for every year, assume the average of the band as the population of the median year t_M of the band, and record this number as the approximate size of birth cohort for year for year t_M years ago. Given that the lion's share of educational enrollment was concentrated in primary school, we suppose that they provide better estimates in alignment with general population trends in India.

We recognize that these numbers are not perfect for a technical measurement of concentration of educational investments since a only fraction of individuals survived until the age of 15 conditional on surviving until age 8. However, these individuals represent the never-takers in enrollment since enrollees are already accounted for in our data. Conditional on not enrolling in primary, these individuals would not have enrolled in higher levels of education. Therefore the calculation does represent rough but meaningful cohort sizes. What admittedly remains debatable is the use of the annual average population for age band of primary going age given that the populations decrease monotonically over age of a cohort, even if our supposition is that alternative specifications is unlikely to change our estimates meaningfully.

The *Progress of Education in India* reports records quinquennial data on enrollment by stages of school education—namely, primary, middle and high school—in addition to enrollment in primary and secondary schools and colleges. However, there are substantial variations in the definition of what constitutes as primary, secondary and high school in the different provinces of British India, some of which persist in the early parts of post-Independence.

3.1.2. Concentration by cohort. Expenditures and enrollment values vary by year at a less natural rate than population. Therefore, we run the risk of under or overestimating concentration by applying the same technique as that of population. For this reason, in this section, we estimate cohort wise educational expenditure.

For simplicity, we begin by taking college as an example and consider the cohorts born in the years t, $t + 1$ and $t + 2$. These cohorts attend primary school in years $[t + 6, t + 10]$,

$[t+7, t+11]$ and $[t+8, t+12]$, respectively, of which only the years $[t+8, t+10]$ overlap for all three cohorts. Likewise, they attend secondary school registered in years $[t+11, t+16]$, $[t+12, t+17]$ and $[t+13, t+18]$, respectively, of which only the years $[t+13, t+16]$ overlap for all three cohorts; they attend college in years $[t+17, t+19]$, $[t+18, t+20]$ and $[t+19, t+21]$, of which all three overlap only in the year $t+19$. Therefore the college enrollment and expenditure data for year $t+19$ allows us to backtrace the expenditure on education for the three generations born between $[t, t+3]$. Due to year on year variations in educational expenses, we only consider the years where the three consecutive cohorts overlap in every level of schooling; substituting by $t' = t+19$, the three ranges of interest in our data for year t' are years $[t'-11, t'-9]$, $[t'-6, t'-9]$ and t' for primary, secondary and college education respectively. For the two overlapping ranges for primary and secondary education, we take the population weighted average expenditure of the three years to get the first estimate of educational expense of the three cohorts. Since the birth cohort t would be represented in the expense calculation for birth cohorts $[t-2, t]$, $[t-1, t+1]$ and $[t, t+2]$, we consider the sum of weighted average of the three calculations as the life-course educational expense for birth cohort t who attend college in year $t+19$.

Let $E^c(t')$ be the total expense on college education and $n^c(t')$ the total college enrollees in year t'. Then

$$\overline{E}^c(t') = \frac{E^c(t')}{n^c(t')} = \frac{E^c(t')}{\sum_i n^c_{t+i}(t')}$$

represents the average educational expense per college student in year t', where $n^c_{t+i}(t')$ is the number of college students registered in data for year t' born in year $t+i$. We know that for $t' = t+19$, only $i = 0, 1, 2$ are permissible. $\overline{E}^c(t'-1)$ and $\overline{E}^c(t'-2)$ are the only other data points where cohort t is represented. We estimate the average college expenditure on generation t as the average

$$\overline{E}^c_t = 3\frac{\sum_{i=17}^{19} n^c(t+i)\,\overline{E}^c(t+i)}{\sum_{i=17}^{19} n^c(t+i)} = 3\frac{\sum_{i=17}^{19} E^c(t+i)}{\sum_{i=17}^{19} n^c(t+i)} \tag{3.1.4}$$

where both $E^c(t+i)$ and $n^c(t+i)$ are observable and 3 represents the length of the program.

More generally, for any educational level j, we estimate the total expense of cohort born in year t as

$$\overline{E}^j_t = \left(t^{\max}_j - t^{\min}_j + 1\right)\frac{\sum_{i=t^{\min}_j}^{t^{\max}_j} E^j(t+i)}{\sum_{i=t^{\min}_j}^{t^{\max}_j} n^j(t+i)} \tag{3.1.5}$$

where $t^{\max}_j$ is the maximum age of enrollment in stage j and $t^{\min}_j$ is the minimum age. For colonial India, we rely on lengths of various states of education on Progress of Education in India reports. Despite variations by province and period of time, the reports from 1870 to 1940 on an average list the following as duration: primary stage as 5 years, middle stage as 3 years, high stage as 3 years and college for 2 years.

3.1.3. Accounting for educational stage data. Unfortunately, expenditure data in PEII are not available at the stage level, but at the institutional level. Enrollment data are available for institutional level for every year of PEII and stage level for some years of PEII. Enrollees in secondary schools include enrollees are primary stage as well as middle stage, but enrollees in primary schools include students (for most part) only in the primary stage.

For the years where stage level data are available, we exploit this difference to derive the approximate number of students who are in primary stage in secondary schools by taking the difference of the number of enrollees in primary stage and enrollees in primary school: if M_i^j is the enrollment of students at stage i in institution j, then $M_p^s = M_p - M_p^p = M_p - M^p$ at a given year. Consequently, we get the number of secondary students in secondary schools as $M_s^s = M^s - M_p^s$. Both middle and high stage students are enrolled only in secondary schools and therefore they add approximately to the number of students are secondary stage M_s^s.

We collect data for institutional level enrollment in British India at an annual level beginning 1870 to 1919 and for years 1927, 1932 and 1937. Stage level data are available only quinquennially beginning 1891. For the years where stage data are not available, we look for the most stable of long run trends in ratios of various enrollments by stage and institutions and use it to interpolate values for those ratios, and then use their linearly interpolated values for years in between to derive further missing variables. The ratio of primary students at primary stage with enrollment in primary institutions, namely $M_p/M^p \approx 1.1$ is highly stable; we use it to derive primary stage enrollment as $M_p = 1.1M^P$. Having derived M_p, we derive secondary enrollment in secondary schools, that is M_s^s. To estimate the enrollment of middle stage students in secondary schools M_m^s, we use the interpolated values of ratio M_m^s/M_s^s. Finally, we estimate the enrollment in high stage as $M_h^s = M^s - M_m^s$.

Some troubles persist for the category of "special schools" since they represent the aggregate of a variety of vocational training schools, specialized instruction, schools for the nobility and so on. Given an over 20 times per year expense for an enrollee of special school in comparison to those enrolled in primary schools, we make the assumption that enrollees in special schools reflect graduates of primary schools who pursue specialized instruction for four years. This assumption is based on the fact that special schools could reflect equivalents to primary school for some groups, and secondary equivalent to others. In sum, we only present results for the assumption of 4 years, although we do recognize its limitations.

For colleges, we aggregate arts and professional colleges into one category, i.e. college. Vocational schools that cater to specialized post secondary level instruction are classified as professional colleges by the PEII reports which we incorporate in a similar fashion. Specific post bachelors data for university students are not separately considered in our estimation unless they are subsumed in arts and professional colleges already in the PEII classification, and therefore our estimates of concentration may be conservative.

Accounting of expenses is done analogous to enrollment. We equal split total expenditure on enrollees of a type of institution by enrollment in the same type of institution to get

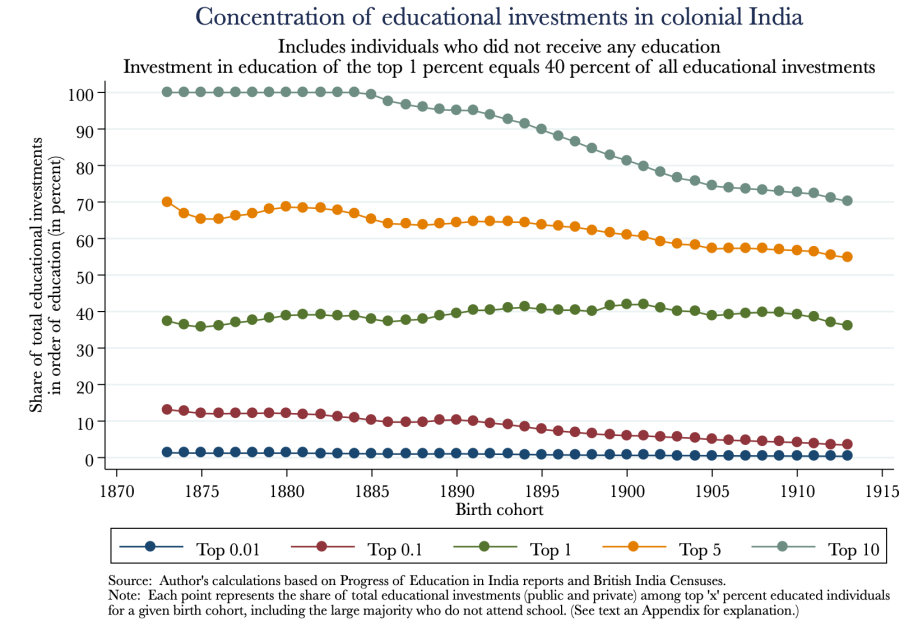
an estimate of per student expense. Despite differences in per student expense of primary school and per student in secondary school (which includes primary level), we assume the latter for primary enrollees in secondary schools and estimate the per student expenditure of students at primary stage as the enrollment weighted average of the two. The investment per primary school student (not per stage) is used for estimating life course investment of special school students since special school students may or may not have conducted their primary schooling in a secondary school.

Colonial concentration. The results of our calculation of concentration of investments are summarized in Figures 3.1.1a and 3.1.1b that split the analysis for the population level concentration and among those enrolled respectively.[1] In line with census estimates for literacy and Barro-Lee data on educational attainment, our results suggest that the bottom 90 percent population never attended schooling until mid 1880s. This period highlights the emergence of schooling for a wider population. Therefore, all investments occurred in the top 10 percent. However, there are sizable inequalities within the top 10 percent, with the top 0.1 percent's share diminished from 10 percent to about 5 for cohorts between 1870 and 1920. However, the share of the top 1 percent remained nearly constant for the full period with concentration of about 40 percent of all investments. The share of the remainder 9 percent diminished as primary enrollment surged for cohorts 1885 onwards. Since the top 1 percent constituted in large part college, special school and high stage enrollees, this suggests that there was increasing preference for investment in tracking forward those who persisted than in ensuring higher investments for the bottom groups.

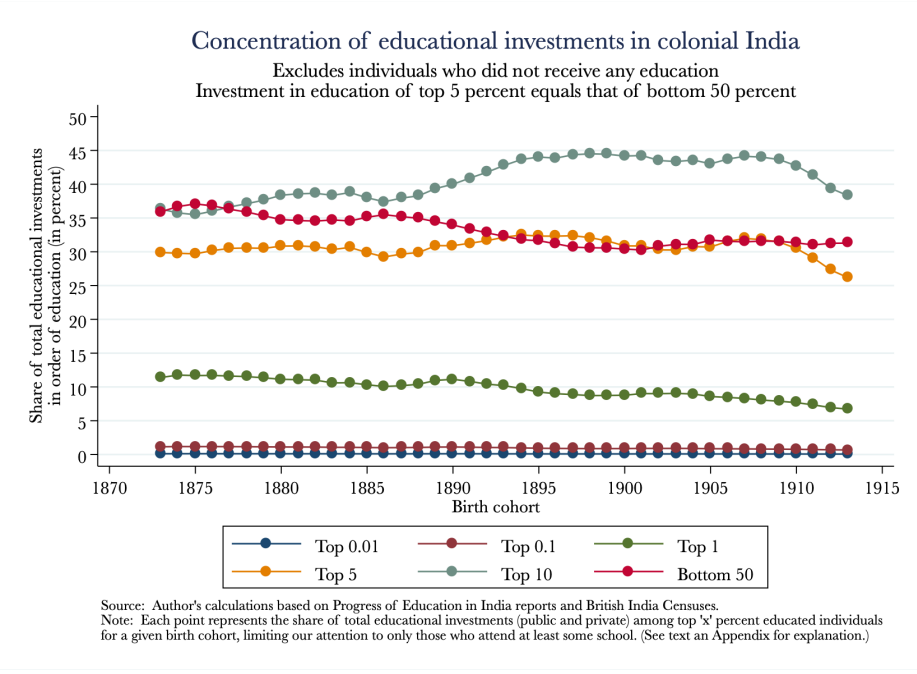
A fuller picture is provided by the concentration of investments among just those who were enrolled (Figure 3.1.b). In line with population estimates, the share of top 1 percent among enrollees declines by 5 percent from 12 to 7 in the nearly forty years of cohorts. However, the share of the top 5 percent (mainly college and high school attains) remained relatively stable and equal to the share of the bottom 50 percent (mostly primary attainees) of all enrollees born in the same year at around 30 to 35 percent suggests that despite a rise in enrollment, preference for investment in those who persisted did not diminish relative to those who newly entered school. The overall stable or increasing share of the top 10 in general over the period above 35 percent supports the claim further.

In all, the results suggest that the inequality in educational investments reduced from 1870 to 1920 for the population due to surge in enrollments and investments at the basic level. However, high and stable concentration of 40 percent in the top 1 percent individuals and 30 percent among the top 5 percent of those who get at least some education suggests that priority has historically been concentrated on those who will go on to pursue higher education, and not on those who have recently gained access to education in the first place and those would eventually drop out of school. If as recently as ten years ago it was only the

[1] See Figure A.0.4 for decomposition of sources of educational expenditures for 1877 to 1920.



(A)



(B)

FIGURE 3.1.1. Evolution of colonial concentration of educational investments

richest few who completed college or secondary school as the previous chapter suggests, then it may be readily inferred that the concentration of educational investments by education also correspond broadly to concentration by income groups in the colonial period. It may indeed be said that the richest few therefore not only had an enormous advantage in accessing human capital, but for any rupee spent, it was mostly they who were advantaged by it.

3.2. Modern concentration: return to the colonial?

For the modern analog, we compare data for about thirty year period between 1986 and 2014. Unlike THE administrative data for colonial India, we use the National Sample Surveys on education and educational expenditure available decennially that capture individual level expenses on education (See the section on data description for construction.) Therefore, we estimate cross sectional rather than cohort based concentration due to absence of annualized or near annualized series on educational expenditure. Further, since only public, and not private, capital expenditure data are available, we restrict our estimation to benefits accrued by households through government support and out of pocket expenses on education. Because our data allows reliable estimation of cross sectional concentration rather over cohort based, we account only for the enrolled students at each period rather than the total population of all individuals who are of school and college going age to avoid difficulties in interpreting concentration that includes a large variation in birth cohorts among those who are not enrolled (or who drop out at some point) in the population. Moreover, household expenditure wise percentiles also help us compare the results on concentration of investments with that of access given they employ the same reference values, even if over different time periods.[2]

Recent expenditures on education demonstrate a high dispersion within levels of schooling; as such, primary school expenditure among the richest households may exceed the college expenditure among the poorest. Therefore accounting for college expenditures before accounting for primary expenditures may not provide precise estimates of concentration in a cross sectional setting.[3] Since unlike the colonial period, recent data allows us to study the full distribution of expenditure on education by household consumption per person, we utilize the same as the basis for understanding concentration of educational investments at each period.

First, we estimate the concentration of educational expenditures by household percentiles at the national level by taking the ratio of aggregate private expenditure on education for the relevant household consumption per capita percentiles and the national private expenditure.

[2]Recall that access was estimated for those who had completed their education by a given year, whereas concentration is estimated for those who are enrolled in the same year. Since life-course schooling lasts roughly for less than 25 years (less than 17 years for the majority), the data on enrollment in 1986 and 1995 (used in the concentration calculation) is included in 2004 and 2011 access data.

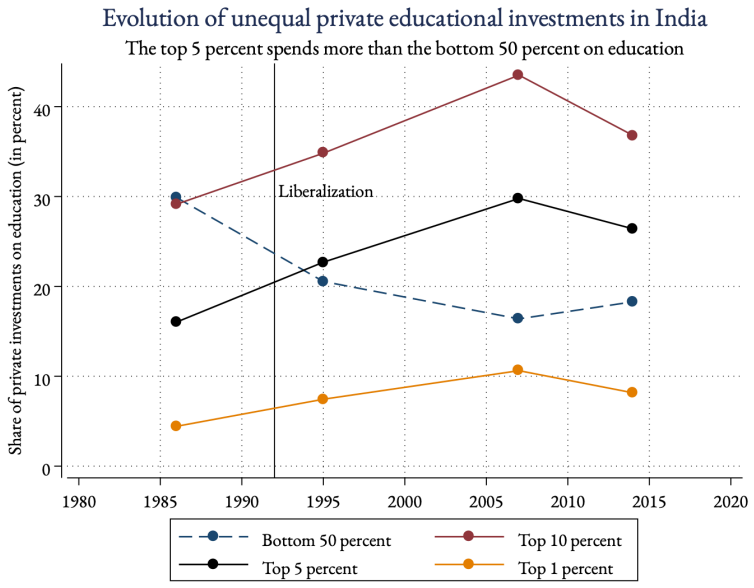
[3]We do not face a similar problem in cohort wise concentration since accounting for college expenses of individuals also allows us to account for primary and secondary school expenditure.

The results are presented in Figures 3.2.1a-b and are straightforward to interpret: the concentration of educational investments among the top at the top level has increased for all groups over the recent thirty years and the share of bottom half of all enrollees has systematically declined. The share of top investments among the 5 percent has nearly doubled over the period and exceeded that of the bottom 50 which declined by over a third from 30 to 18 percent. Furthermore, the stability in bottom 50 since 2005 and the decline in the top 10 percent, suggests an increasing concentration in the 'next 40' (top 10 to top 50) that form the bulk of the middle class of India. Disproportionate expenditure by the top groups suggests an increasing role of private finance in accruing human capital and rising income inequality suggests that this may create further unequal opportunities should education privatize further.

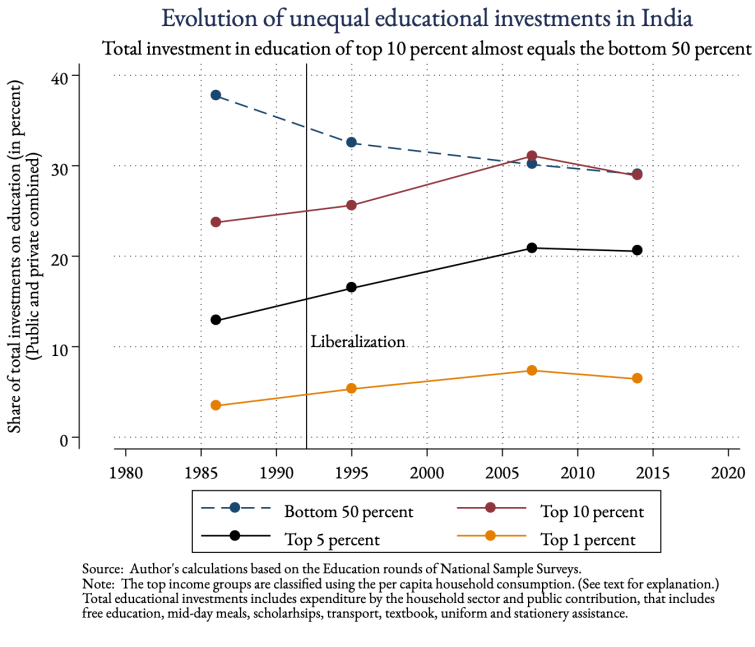
Nevertheless, concentration by private educational expenditures misrepresents the true inequality since a majority of the bottom 50 percent access public or public-private partnership schools that charge low to no fees, and receive government support in scholarships, mid day meals, textbook subsidy, stationery travel concessions and tuition waivers. We account for these factors by estimating the total educational investments combining the public assistance to individuals on each component of education related expenditure and the out of pocket household expenditures.[4] Having estimated public contribution for each individual, we estimate the total educational investments as the sum of public and private (household) expenditures. The concentration is estimated in a similar manner as the private, and the results are presented in Figure 3.2.1b. First, although the total investment concentration is more optimistic than private, it follows the same patterns: a decreasing share of the bottom 50 from over 36 to 30 percent, and an increasing share of the top 5 from about 12 to over 20 percent. The share of the top 10 enrollees equals almost that of the bottom half of all enrollees and is overrepresented by more than four times. Similarly, the top 1 percent's expenditure has grown from about five to over seven times their share of enrollment. Second, the share of the 'next 40' has remained stable with a change over around 3 percent due to near equal displacement of the top 10 and bottom 50. This suggests that the bulk of the middle has gained equally from public funds as much as it spends out of pocket to access education over time.

The results suggest that the increase of concentration of educational investments has been systematic over the three decades regardless of public or private spending. Despite a generous estimation of public contribution, the patterns do not vary in private and total investments over the years, nor are they affected by the campaigns to universalize education between 1995 and 2005. Between 2005 and 2014, although both public and private concentration declined among the top 10 percent, the share of bottom 50 declined in total investments continued to decline, suggesting increased access and expenditure among the middle class. Whether this was a result of the Right to Education Act of 2009 or not deserves further exploration.

[4]See data description for details on accounting for public contribution.



(A) Private investments



(B) Total investments

FIGURE 3.2.1. Evolution of recent concentration of educational investments

3.2.1. Variation in recent concentration by ideology. The depressed concentration when accounting for public contribution suggests that states where large shares of households benefit from government support are likely to have more equitable educational investments than those where privatization is more more prevalent. To test if this is the case, we explore the differences in concentration by the same nine states of various roughly defined ideologies as the one in educational access. Figures 3.2.2 and 3.2.3 describe the estimated inequality in private spending and total investments by states. For clarity, we only present data for 1986 and 2014, and, for comparability, we use the national household consumption per capita percentiles. Tables B.0.5 and B.0.6 of Appendix summarize the same results but for a greater variety of states.

First, regardless of private or total spending, inequality has by and large increased for every state in our analysis with increasing share of the top fractiles and decreasing share of the bottom fractiles. The change over thirty years indicates a near substitution of shares between the top 10 and bottom 50 among private investments; this implies the the share of the middle class (next 40) has remained relatively stable in contrast to the bottom and top groups. This is less obvious in total investments where the change in middle class's share have increased in some and decreased in others. This variation may reflect state specific policies in dealing with expansion of schooling over the period.

Second, differences by ideology for educational investments are absent in contrast to educational access in both magnitude and pattern: concentration has increased regardless. States such as Kerala and Gujarat show strikingly similar patterns in both private and total investments in 1986 and the change up to 2014 for each major fractile, as do say Tamil Nadu and Madhya Pradesh. Although the data excludes capital investments for both public and private sectors—which may show variation between left and right leaning states for the bottom groups—the similarities in distributional private and total spending (that accounts for effective public assistance including free education) are difficult to overlook. This suggests that despite closing gaps in educational access between privileged and underprivileged minorities, i.e. the social divide, class divides in both access and investment are significant.

Third, the gap between total and private shares of investments for the bottom 50 and the top 10 have increased on an average over the thirty years. Figure 3.2.4 highlights the change over the thirty year period for all major state regions covered in Tables B.0.5 and B.0.6. By accounting for public contribution, the relevance of government has increased over time in reducing inequality of human capital investments, particularly for the bottom half of all enrollees.

The results on recent concentration, both nationally and by states, suggest that opportunity has become more unequal since fewer resources are being allocated to the bottom groups in contrast to increased concentration for the top groups by their level of family income or consumption. Our results may be more disappointing once we account for the fact

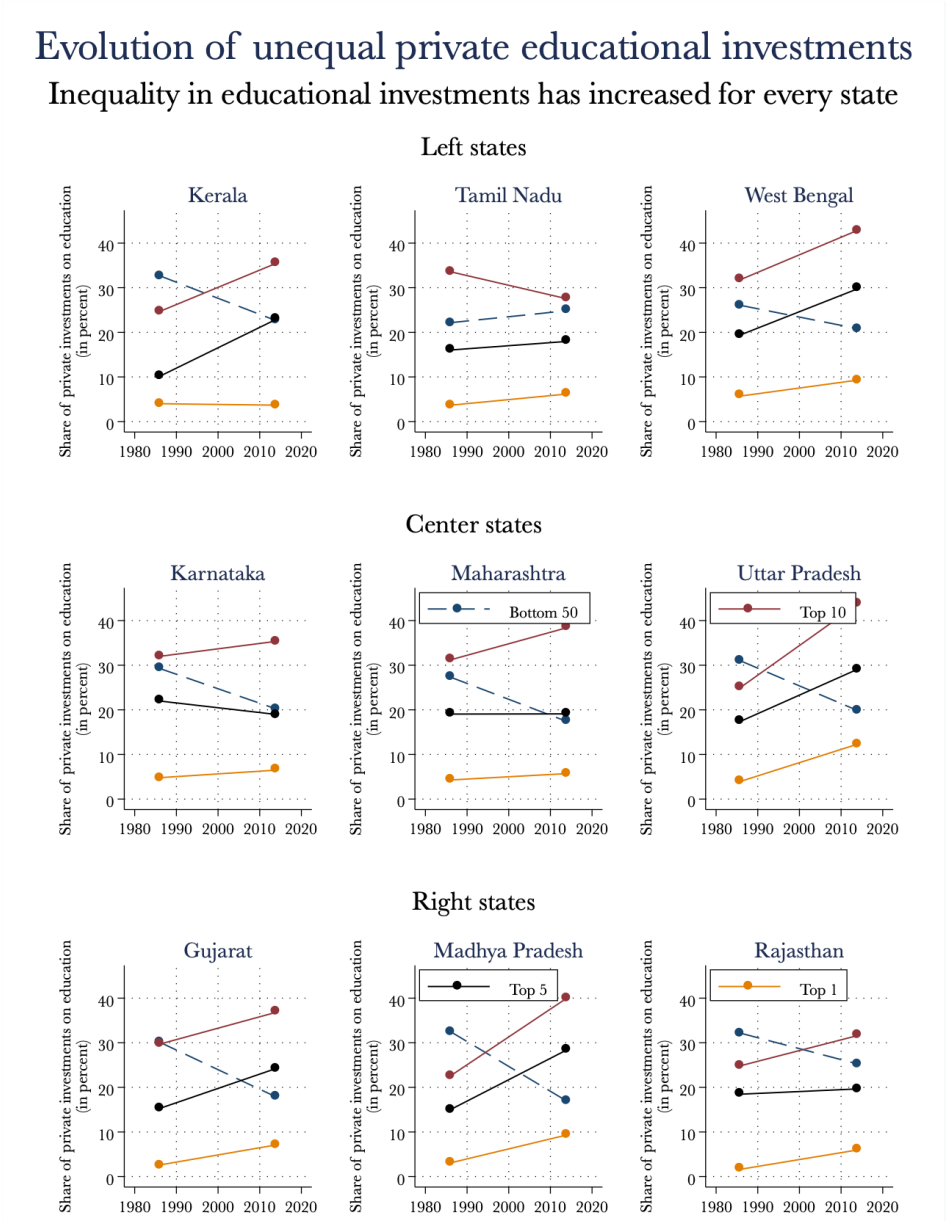


FIGURE 3.2.2. Recent concentration of private educational investments by state

that enrollment at top levels has been higher historically and that the bottom groups see a surge in the recent period due to policies on universalization since government spending has not increased at the same pace. In fact, recent developments indicate a greater involvement

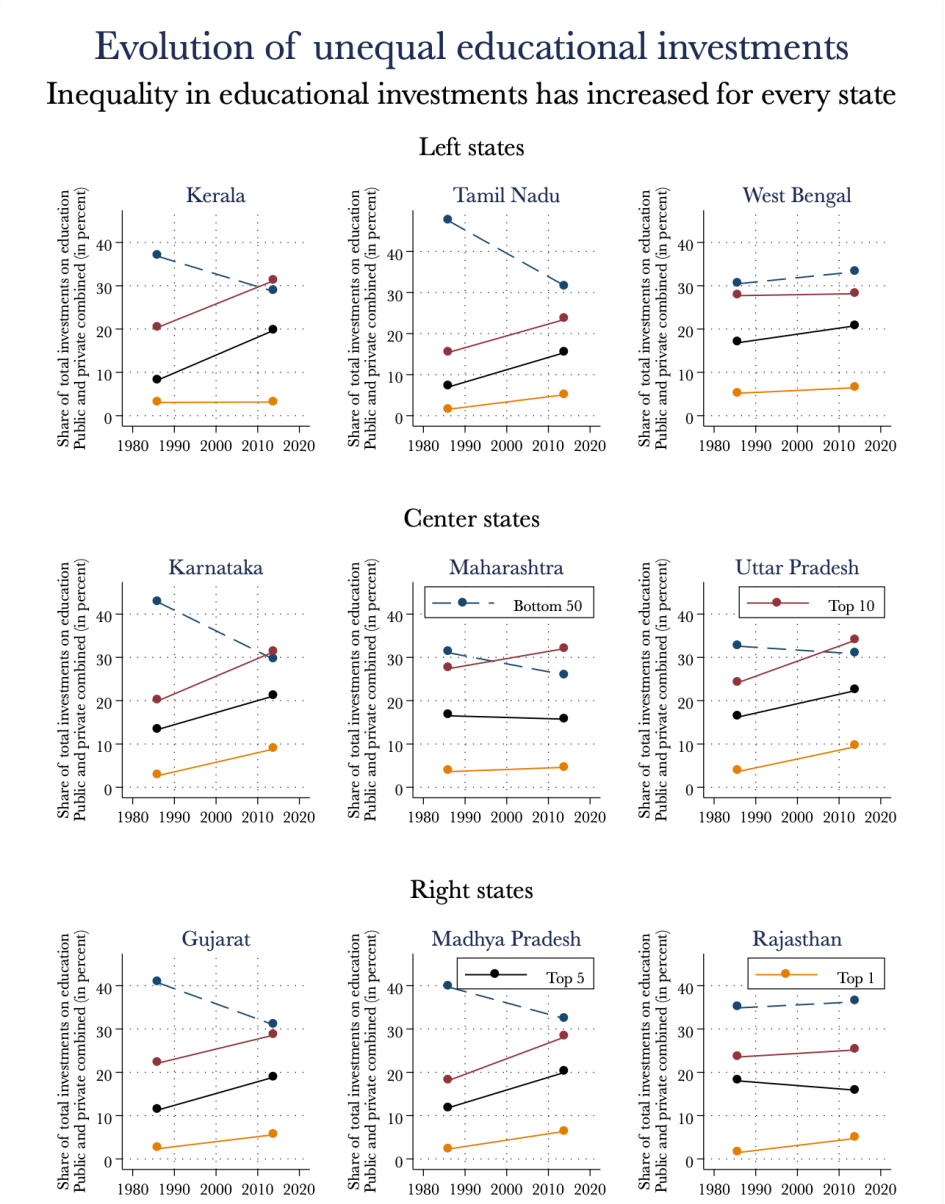
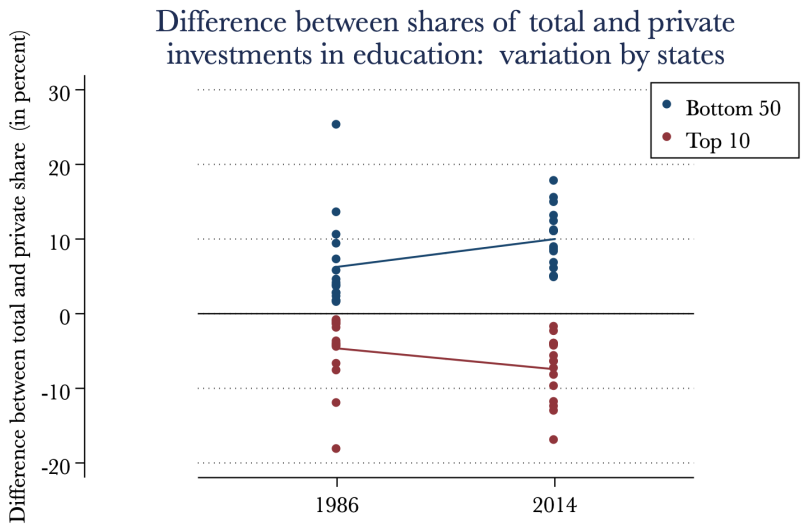


FIGURE 3.2.3. Recent concentration of total educational investments by state

of private players in educational sector and the New Education Policy of 2020 only encourages it further. Reflecting on the share of top 10 and bottom 50 percent, it is easy to notice that their equality of shares was held for cohorts of 1870s among all enrollees as it does as



Source: Author's calculations using the Education rounds of National Sample Surveys.
Note: Each point represents the difference in shares for a state for the given year. Bottom 50 and top 10 percentiles refer to national percentiles of household per capita consumption. The straight lines are linear fits over the period for illustrative purposes: states are unweighted by enrollment or expenditure.

FIGURE 3.2.4. Role of public investments in reducing concentration

recently as 2014. With increased spending for the top groups, one wonders if we were to return to colonial levels of concentration in human capital of those who pursue education, granting that the share of those who never enroll in school has declined dramatically. The policy implications of the results on recent period are clear: since public funds play an increasing role in bridging the concentration of educational investments, educational budgets of the central and state governments need to expand to compensate for increasingly unequal access to opportunity.

3.3. Summarizing concentration

In this part of the thesis, we first develop a way to estimate both cross sectional and cohort based concentration of educational investments using archival administrative data and representative surveys. We focus on two periods in Indian history: 1870 to 1915 when formal schooling emerged in colonial India, and 1980 to 2015 when education went from being a privilege to a human right. Our cohort based estimates for the colonial period show that over 90 percent did not access school until 1880s, and despite surge in enrollments since, the concentration among the top 5 percent enrolled has been constant around 40 percent and the share of the top 10 percent always exceeded 35 percent. The share of the bottom 50 percent reduced from over 35 to about 30 percent in the forty year period between 1870s 1910s. Recent cross sectional evidence for the thirty years after 1980s suggests that the share

of the bottom 50 has declined from over 35 percent to about 30 percent, a pattern replicated after a century, and the share of top 5 jumped from about 12 to over 20 percent. State by state variation suggests that long run ideological differences in party rule may not translate to concentration of educational investments the same way they did for access, and that inequality of investments has increased nearly universally. The increasing gap in bottom shares when accounting for public funds with respect to private expenditure suggests that government financing plays a major equalizing role in ensuring students receive sufficient access. With recent developments towards privatization of education, our results suggest that these may increase inequality of opportunity.

The results are not full-proof and there are a few limitations. Conceptually, concentration of educational investments per se does not capture differences in quality. A state for instance may simply be more effective at choosing avenues of expenditure to produce higher quality than another state that spends substantially on the bottom 50. Moreover, an argument may be made that expenditure does not translate to quality. Indeed, an increased private educational expenditure for the rich may simply reflect market adjusting its prices for the same good, where the consumer pays for discriminating against co-enrolling with the bottom groups. While we do believe that quality-adjusted cost estimations are essential for capturing true levels of inequality, we provide two defense to our analysis. First, an additional unit of money may not be marginal in the context of India since the starting point is already too low. Aggregate financing of education is likely to boost wages, attract higher quality teaching staff and provide an overall ecosystem of infrastructure that may be more conducive to learning. That said, it remains essential to determine the effectiveness of funds and choose the optimal ones. Randomized trials of such type include. Second, given absence of literature in this domain, our estimates are only a starting point for further analysis.

Furthermore, the colonial calculations do not decompose unit expenditures by grade of study during school. Our section on access demonstrates an incredible drop out rate grade 1 or 2 for instance as late as 1930s. Naturally, the per enrollee expenditure will be low for the first two grades, and higher for the latter. Our estimation an averages out expenditures for every grade of school. A further technical note concerns our use of nominal values of currency rather than inflation adjusted values over the years. For instance, in accounting for expenditure during primary school of an individual who studied college, the nominal values may overstate the life-course expenditure since college expenditures values will reflect the real value plus the inflation over time. This may be adjusted for in further work.

In the recent period, we only account for gains and expenditure by the household in the recent period and not capital or institutional expenditure by the public and private sectors. A large part of the reason is lack of a reliable data series on private capital expenditure and its beneficiaries. Our data on the population of schools and universities in India may be utilized in further work in this direction, although we reckon that accounting for fixed capital will be

difficult. That said, incorporation of institutional and capital spending may not provide more optimistic estimates since quality adjusted unit labor costs may be similar for both private and public, and capital expenditure per enrollee of the top 1 percent is likely to be higher than capital expenditure per enrollee by the public sector among the bottom 50. Therefore, at best, estimates of inequality may either look similar or more skewed in favor the richest. Moreover, sample surveys do not capture the incomes and expenditures within the top 1 percent, which tends to be sizably higher, which may further skew our results and be more pessimistic than the colonial period. Further work will benefit from incorporating the same to provide true levels of inequality.

Summing up

The evidence in this thesis highlights the broad cleavages in educational access and investments for two key parts of Indian history: when formal schooling emerged during the colonial period, and when education became universalized during the turn of the millennium. Although aggregate attainment has increased, we find expected gaps by sex, region and caste. But more importantly, we find that Brahmins and Muslims—by way of saying privileged minorities and underprivileged minorities—demonstrate a persistent gap in access both over the course of the century at every level of schooling and family income. Patterns of enrollment by level of schooling are replicated over time. While these gaps are closing among the richest households in the last couple of decades, they remain substantial. We find some evidence that differences between two types of minorities may be driven by differences between states by ideology, however, the concentration of investments by class suggests that these differences may be attributable to class gaps. Moreover, gains in education have been accrued by the richest households in the decades that followed Indian Independence despite having a head-start in pre and post secondary schooling.

Investments remain concentrated among the top 1 to 10 percent. During the emergence of schooling in colonial period (1880s to 1910s), nearly an equal share of investments of about 30-35 percent were made among the top 5-10 and bottom 50 percent enrollees. In the recent period where education universalized (from 1980s to 2010s), investments in the bottom 50 percent households has declined over time in the recent period to equal the top 10 percent. These trends are worrying since they do not capture the top level inequality among the 1 percent (due to lack of representativeness in surveys for the richest households) and institutional expenditures by public and private sectors in education (due to lack of data); accounting for them is likely to show a further increase. Indeed, our findings are subject to data limitations, and we fully acknowledge them. Our evidence suggests that we need a broader research agenda to harmonize long run data to better understand inequality of opportunity. This thesis is only a step in that direction.

As an implied conjecture, we believe that systematic inequality of opportunity between the two sets of minorities may explain a part of long run divergences of political preferences. Indeed, differences in how communities view their identity may influence their behavior to reinforce their experiences through self-fulfilling prophecies.(Singh and Husain, 2016) Since ethnic networks play an important role in enhancing opportunity for members of respective

communities, it is vital that efforts are made to bridge—and not polarize—the privileged minorities from the underprivileged minorities through new identity channels to improve educational and labor market mobility.(see Munshi, 2019)

The policy implications of our findings are that the state will benefit by increasing public spending on education—it plays an even greater role today than it did once in reducing inequality of investments. Enhancing education at the bottom of the income distribution will be essential to catch up with recent developments in skill biased technological changes. Indeed, these will serve as investments in the long run, and reduce the growing ethnic tensions unfortunately faced by the country today. Although the long arc of history may eventually yield the Kuznets curve in education (as Morrisson and Murtin, 2013, data may suggest), there is no reason to believe that this will be automatic or soon: policy makers are best informed to make decisions today to ensure that India does not continue to remain a society that rewards the accident of birth.

Bibliography

Aghion, P., B. F. Jones, and C. I. Jones (2017). Artificial intelligence and economic growth. Technical report, National Bureau of Economic Research.

Alesina, A., R. Baqir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly journal of economics 114*(4), 1243–1284.

Ambedkar, B. R. (1936). Annihilation of caste.

Asadullah, M. N. and G. Yalonetzky (2012). Inequality of educational opportunity in india: Changes over time and across states. *World Development 40*(6), 1151–1163.

ASER (2010). Annual survey of education report 2010. Technical report, Pratham.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics 118*(4), 1279–1333.

Banerjee, A., A. Gethin, and T. Piketty (2019). Growing cleavages in india? evidence from the changing structure of electorates, 1962–2014. *Economic and Political Weekly 54*(11), 43–44.

Barro, R. J. (2001). Human capital and growth. *American economic review 91*(2), 12–17.

Barro, R. J. and J.-W. Lee (1993). International comparisons of educational attainment. Technical report, National Bureau of Economic Research.

Barro, R. J. and J. W. Lee (1996). International measures of schooling years and schooling quality. *The American Economic Review 86*(2), 218–223.

Basu, A. (1974). *The Growth of Education and Political Development in India 1898-1920*. Oxford University Press: New Delhi.

Basu, B. D. (1867). *History of education in India under the rule of the East India Company*. Calcutta: Modern Review Office [1922].

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy 70*(5, Part 2), 9–49.

Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press.

Benhabib, J. and M. M. Spiegel (1994). The role of human capital in economic development evidence from aggregate cross-country data. *Journal of Monetary economics 34*(2), 143–173.

Bharti, N. (2018). Wealth inequality, class and caste in india, 1961-2012. *Paris School of Economics, Master Thesis*.

Bonneau, C. (2019). The concentration of educational investment in the us (1970-2018), with a comparison to france. *Paris School of Economics, Master Thesis*.

Borjas, G. J. (1992). Ethnic capital and intergenerational mobility. *The Quarterly journal of economics 107*(1), 123–150.

Borooah, V. K. and S. Iyer (2005). Vidya, veda, and varna: The influence of religion and caste on education in rural india. *The Journal of Development Studies 41*(8), 1369–1404.

Bratberg, E., J. Davis, B. Mazumder, M. Nybom, D. D. Schnitzlein, and K. Vaage (2017). A comparison of intergenerational mobility curves in germany, norway, sweden, and the us. *The Scandinavian Journal of Economics 119*(1), 72–101.

Brunori, P., F. H. Ferreira, and V. Peragine (2013). Inequality of opportunity, income inequality, and economic mobility: Some international comparisons. In *Getting development right*, pp. 85–115. Springer.

Castello-Climent, A., L. Chaudhary, and A. Mukhopadhyay (2018). Higher education and prosperity: From catholic missionaries to luminosity in india. *The Economic Journal 128*(616), 3039–3075.

Castelló-Climent, A., L. Chaudhary, and A. Mukhopadhyay (2018). Higher education and prosperity: From catholic missionaries to luminosity in india. *The Economic Journal 128*(616), 3039–3075.

Chancel, L. and T. Piketty (2019). Indian income inequality, 1922-2015: From british raj to billionaire raj? *Review of Income and Wealth 65*, S33–S62.

Chaudhary, L. (2006). Social divisions and public goods provision: Evidence from colonial india. *Unpublished Working Paper, Stanford University*.

Chaudhary, L. (2009). Determinants of primary schooling in british india. *The Journal of Economic History*, 269–302.

Chaudhary, L. (2010). Taxation and educational development: Evidence from british india. *Explorations in Economic History 47*(3), 279–293.

Chaudhary, L. (2015). Caste, colonialism and schooling. *A New Economic History of Colonial India*, 161.

Chaudhary, L. and M. Garg (2015). Does history matter? colonial education investments in i ndia. *The Economic History Review 68*(3), 937–961.

Chaudhary, L. and J. Rubin (2011). Reading, writing, and religion: Institutions and human capital formation. *Journal of Comparative Economics 39*(1), 17–33.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics 129*(4), 1553–1623.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives 27*(3), 79–102.

Council of India, B. E. I. C. (1835). English education act 1835.

Davis, L. E., L. E. Davis, R. A. Huttenback, and S. G. Davis (1986). *Mammon and the pursuit of Empire: The political economy of British imperialism, 1860-1912*. Cambridge University Press.

Desai, S. and V. Kulkarni (2008). Changing educational inequalities in india in the context of affirmative action. *Demography 45*(2), 245–270.

Deutscher, N. and B. Mazumder (2019). Intergenerational mobility in australia: national and regional estimates using administrative data.

Dreze, J. and A. Sen (1999). India: Economic development and social opportunity. *OUP Catalogue*.

Easterlin, R. A. (1981). Why isn't the whole world developed? *The Journal of Economic History 41*(1), 1–17.

Engineer, A. A. (2001). Muslims and education. *Economic and Political Weekly 36*, 3221–3221.

Fasih, T., G. Kingdon, H. A. Patrinos, C. Sakellariou, and M. Soderbom (2012). *Heterogeneous returns to education in the labor market*. The World Bank.

French, R. and G. Kingdon (2010). The relative effectiveness of private and government schools in rural india: Evidence from aser data. *London: Institute of Education*.

Fulford, S. (2014). Returns to education in india. *World Development 59*, 434–450.

Giovanni, J. d. and A. Matsumoto (2011). The value of human capital wealth.

Goldin, C. D. and L. F. Katz (2009). *The race between education and technology*. Harvard University Press.

Gounden, A. N. (1967). Investment in education in india. *Journal of Human Resources*, 347–358.

Green, A. (1997). Education and state formation in europe and asia. In *Education, Globalization and the Nation State*, pp. 29–51. Springer.

Green, A. (2013). *Education and state formation: Europe, East Asia and the USA*. Springer.

Hanushek, E. A. (1981). Throwing money at schools. *Journal of policy analysis and management 1*(1), 19–41.

Hanushek, E. A. (2013). Economic growth in developing countries: The role of human capital. *Economics of Education Review 37*, 204–212.

Hanushek, E. A. and D. D. Kimko (2000). Schooling, labor-force quality, and the growth of nations. *American economic review 90*(5), 1184–1208.

Hanushek, E. A. and L. Woessmann (2010). Education and economic growth. *Economics of education*, 60–67.

Harma, J. (2011). Low cost private schooling in india: Is it pro poor and equitable? *International journal of educational development 31*(4), 350–356.

Heckman, J. J., A. B. Krueger, et al. (2005). Inequality in america: What role for human capital policies? *MIT Press Books 1*.

Irving, M. and G. G. Kingdon (2008). Gender patterns in household health expenditure allocation: A study of south africa. *CSAE Working Paper Series 2008 32*.

Kingdon, G. and M. Muzammil (2018). Per pupil expenditure in the government schools of uttar pradesh, and the rate of reimbursement to private schools under the right to education act: An update.

Kingdon, G. G. (2007). The progress of school education in india. *Oxford Review of Economic Policy 23*(2), 168–195.

Kingdon, G. G. (2020). The private schooling phenomenon in india: A review. *The Journal of Development Studies*, 1–23.

Kingdon, G. G. and N. Theopold (2006). Do returns to education matter to schooling participation? *Education Economics 16*(4).

Kothari, D. S. (1970). Kothari commission-educational and national development: Report of the education commission 1964-1966. Technical report, National Council of Educational Research and Training.

Krueger, A. (2012). The rise and consequences of inequality. *Presentation made to the Center for American Progress, January 12th*.

Lee, C.-I. and G. Solon (2009). Trends in intergenerational income mobility. *The Review of Economics and Statistics 91*(4), 766–772.

Lindert, P. H. (2004). *Growing public: Volume 1, the story: Social spending and economic growth since the eighteenth century*, Volume 1. Cambridge University Press.

Lucas Jr, R. E. (2015). Human capital and growth. *American Economic Review 105*(5), 85–88.

Macaulay, T. B. (1835). Minute on indian education.

Mandal, B. P. (1971). Report of the backward classes commission. Technical report, Government of India.

Mankiw, N. G., D. Romer, and D. N. Weil (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics 107*(2), 407–437.

Manuelli, R. E. and A. Seshadri (2014). Human capital and the wealth of nations. *American economic review 104*(9), 2736–62.

McAlpin, M. B. (1983). Famines, epidemics, and population growth: the case of india. *The Journal of interdisciplinary history 14*(2), 351–366.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of political economy 66*(4), 281–302.

Morrisson, C. and F. Murtin (2013). The kuznets curve of human capital inequality: 1870–2010. *The Journal of Economic Inequality 11*(3), 283–301.

Mukherjee, S. B. (1976). *Age distribution of the Indian population: a reconstruction for the states and territories, 1881-1961*. Honolulu, HI: East-West Center, East-West Population Institute.

Munshi, K. (2019). Caste and the indian economy. *Journal of Economic Literature 57*(4), 781–834.

Nurullah, S. and P. J. Naik (1943). *History of Education in India during the British period*. Macmillan, Bombay.

Parliament of India, G. o. I. (2009). Right of children to free and compulsory education act or right to education act (rte).

Patrinos, A. H. and G. Psacharopoulos (2002). *Returns to investment in education: a further update*. The World Bank.

Patrinos, H. A., C. Ridao-Cano, and C. Sakellariou (2009). A note on schooling and wage inequality in the public and private sector. *Empirical Economics 37*(2), 383–392.

Pieters, J. (2011). Education and household inequality change: a decomposition analysis for india. *Journal of Development Studies 47*(12), 1909–1924.

Piketty, T. (2014). *Capital in the 21st century*. Harvard University Press.

Piketty, T. (2015). About capital in the twenty-first century. *American Economic Review 105*(5), 48–53.

Piketty, T. (2020). *Capital and ideology*. Harvard University Press.

Psacharopoulos, G. (1981). Returns to education: An updated international comparison. *Comparative education 17*(3), 321–341.

Psacharopoulos, G. (1985). Returns to education: A further international update and implications. *Journal of Human resources*, 583–604.

Psacharopoulos, G. (1989). Time trends of the returns to education: Cross-national evidence. *Economics of Education Review 8*(3), 225–231.

Psacharopoulos, G. and H. A. Patrinos (2018). *Returns to investment in education: a decennial review of the global literature*. The World Bank.

Ramachandran, V., V. Rawal, and M. Swaminathan (1997). Investment gaps in primary education: a statewise study. *Economic and Political Weekly*, 39–45.

Sachar, R. (2006). Social, economic and educational status of the muslim community of india. Technical report, Cabinet Secretariat, Government of India.

Sen, A. (1985). *Commodities and capabilities*. Amsterdam: North-Holland.

Singh, M. and Z. Husain (2016). Self-fulfilling equilibrium and social disparities in urban india. *Economic and Political Weekly 51*(48).

Tilak, J. B. (1988). Costs of education in india. *International Journal of Educational Development 8*(1), 25–42.

Tilak, J. B. (1997a). Five decades of underinvestment in education. *Economic and Political Weekly*, 2239–2241.

Tilak, J. B. (1997b). Investment gaps in primary education. *Economic and Political Weekly*, 972–975.

Tilak, J. B. (2006). On allocating 6 per cent of gdp to education. *Economic and Political Weekly*, 613–618.

Tilak, J. B. (2007). Knowledge commission and higher education. *Economic and Political Weekly*, 630–633.

Tilak, J. B. (2010). Neither vision nor policy for education. *Economic and Political Weekly*, 60–64.

Tukey, J. W. (1977). *Exploratory data analysis*, Volume 2. Reading, MA.

Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician 34*(1), 23–25.

Whitehead, C. (2005). The historiography of british imperial education policy, part i: India. *History of Education 34*(3), 315–329.

APPENDIX A

Figures

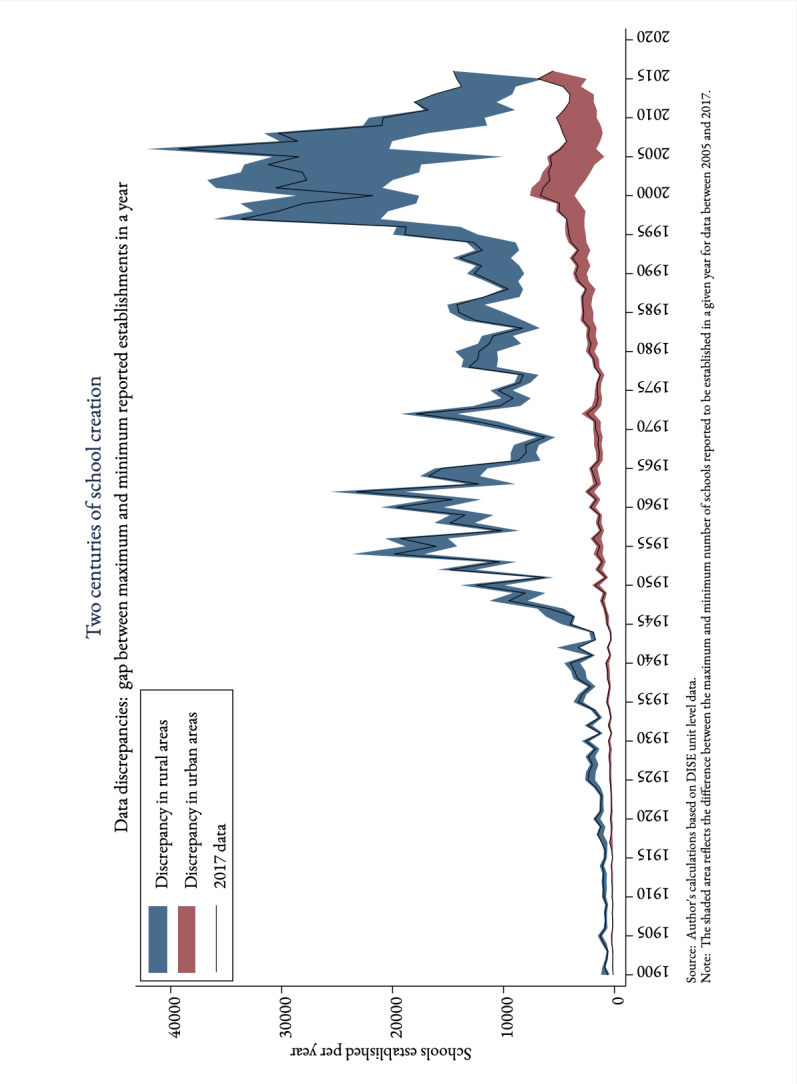
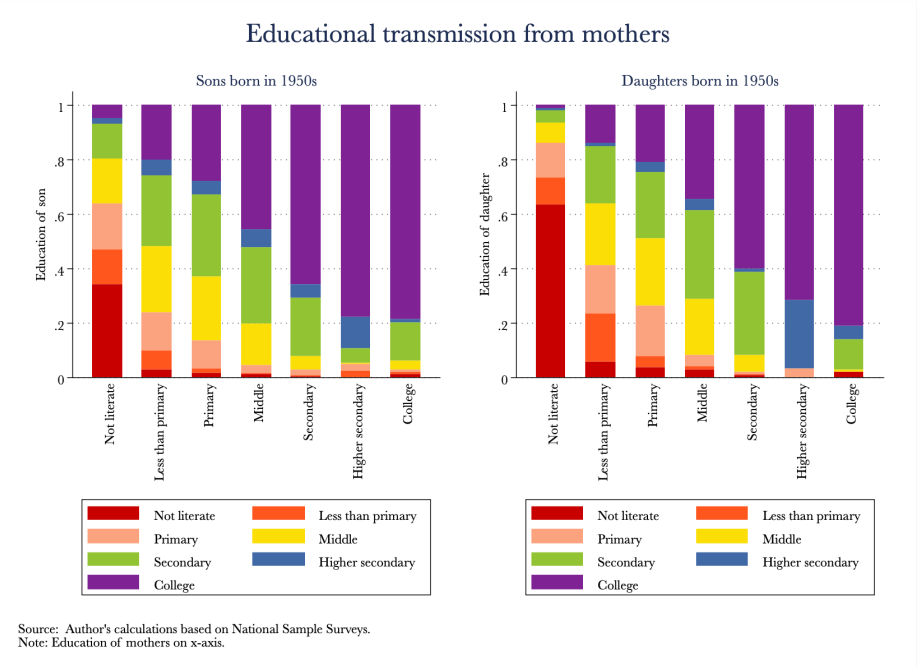
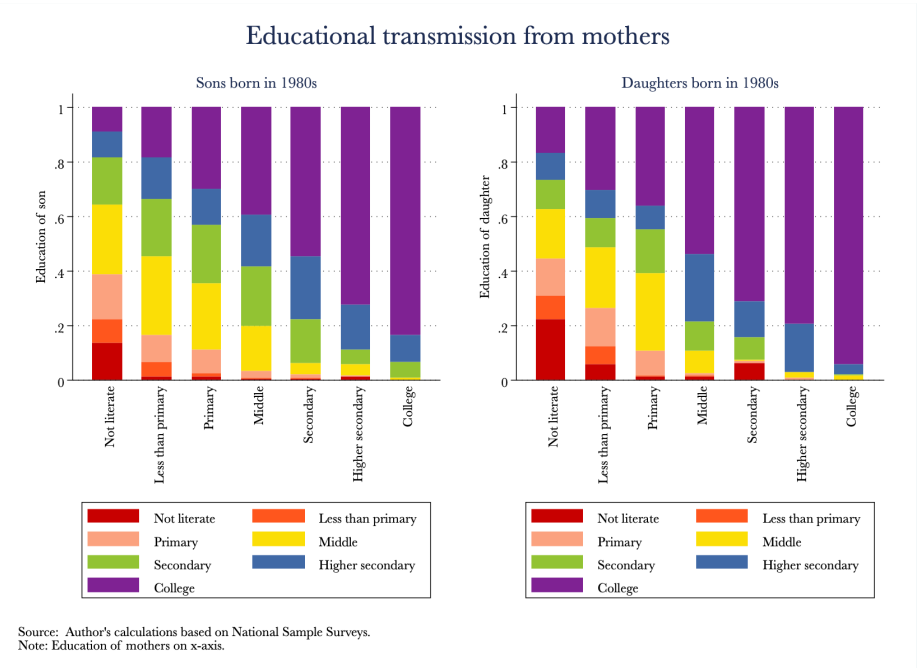


FIGURE A.0.1. Discrepancy in DISE data on school establishment

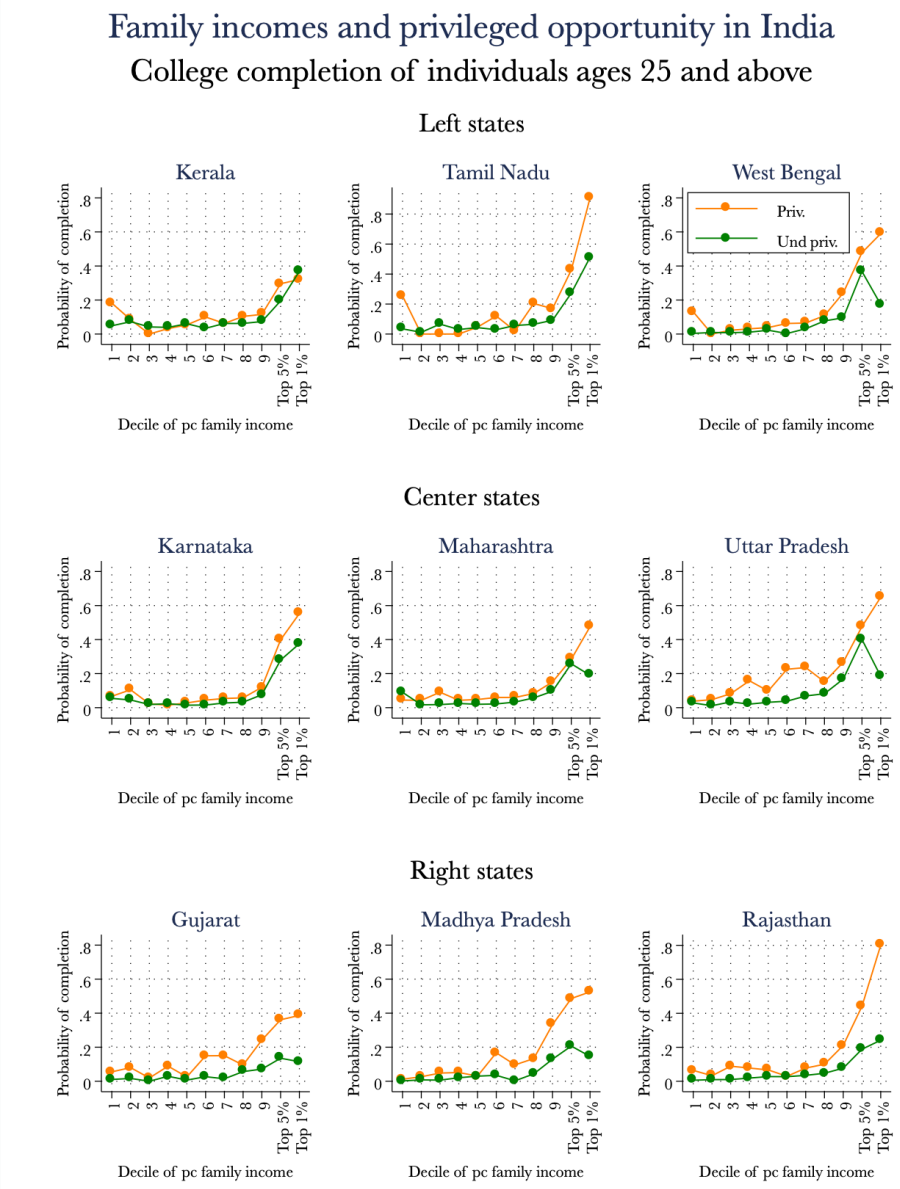


(A) Children born in 1950s



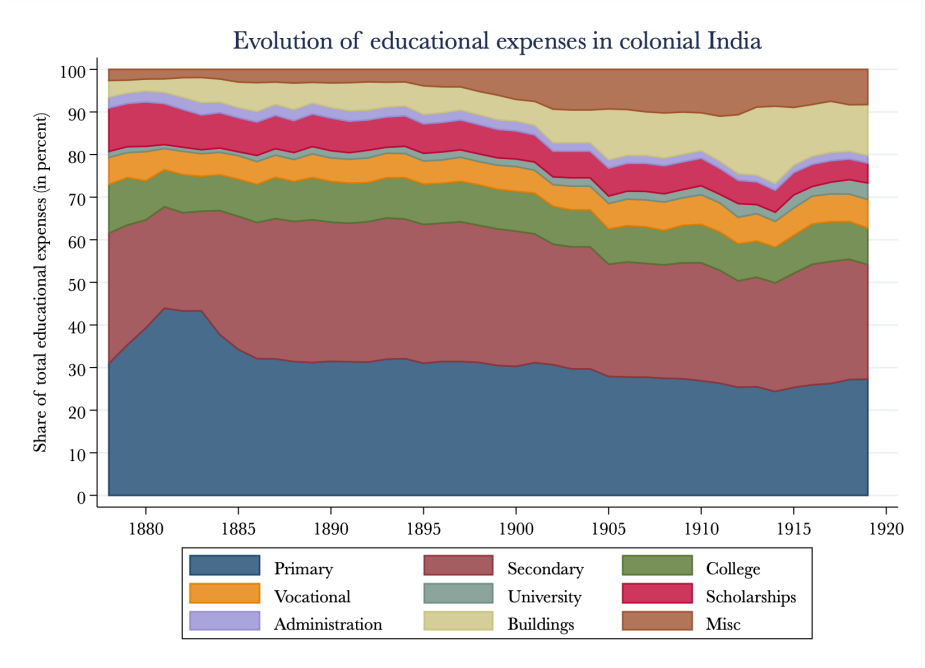
(B) Children born in 1980s

FIGURE A.0.2. Educational transmission from mothers

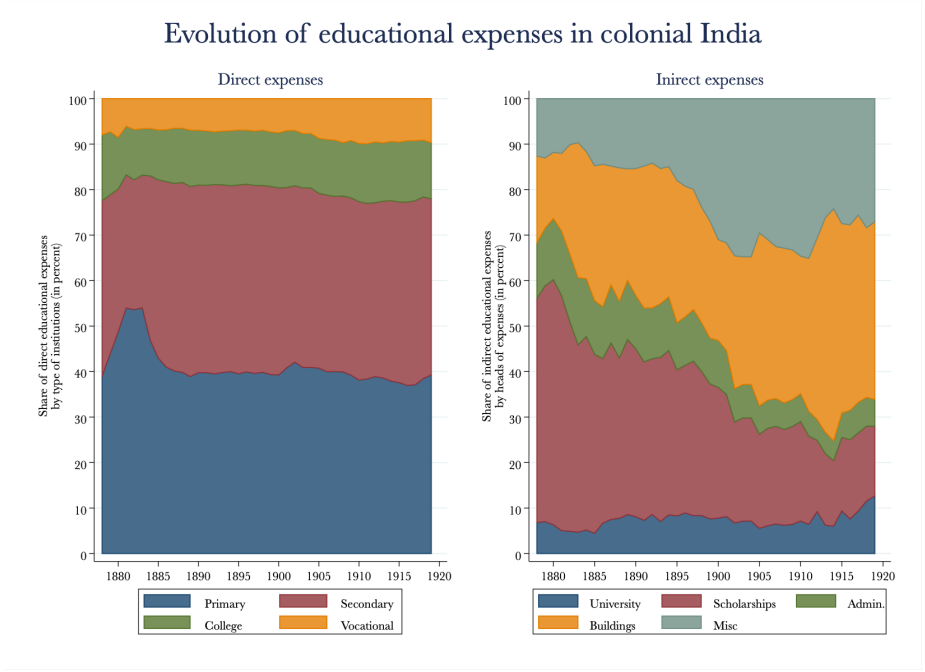


Source: Author's calculations based on ICPSR IHDS 2011.
Note: College completion probability represents the share of percentile population who have a Bachelors degree or at least 15 years of schooling. privileged minorities consists of forward castes, Brahmins, Christian, Jains and Sikhs. Underpriviledged minorities consists of Muslims, OBCs, SCs and STs. See text for explanation on separation of states by ideology.

FIGURE A.0.3. Differences in college completion by family income and states



(A) Decomposition of colonial investments in education



(B) Decomposition by direct and indirect educational investments

FIGURE A.0.4. Shares of educational spending by type of expense

APPENDIX B

Tables

	Enrollment share of population by educational institutions (in percent)					
Year	Primary	Secondary	Secondary and special	Total school	Total college	Total
1931	30.3	3.9	4.6	21.4	0.68	16.5
1926	25.7	3.0	3.7	17.8	0.67	14.6
1921	18.8	2.1	2.1	12.6	0.49	10.9
1916	18.2	1.6	1.8	11.4	0.38	10.3
1911	16.2	1.4	1.7	9.8	0.30	9.2
1906	13.1	1.0	1.2	7.7	0.20	7.2
1901	11.0	0.9	0.9	6.5	0.19	5.9
1896	10.9	0.7	0.7	6.2	0.15	5.9

TABLE B.0.1. Enrollment share by institution type in colonial India

Year	Share of students by stage of education (in percent)									Total students (in thousands)	Total individuals from ages 5 to 18 (in thousands)
	High	Middle	Primary	General school	Special school	Total school	General college	Profess. college	Total college		
1931	2.85	8.09	86.03	96.96	2.24	99.20	0.65	0.15	0.80	12,121	73,249
1926	2.64	6.78	86.63	96.05	3.12	99.16	0.66	0.17	0.84	10,528	71,999
1921	2.83	5.61	89.09	97.52	1.72	99.24	0.58	0.18	0.76	7,742	70,749
1916	3.04	4.02	90.25	97.31	2.03	99.34	0.51	0.16	0.66	7,096	68,694
1911	2.32	4.50	89.67	96.49	2.91	99.40	0.49	0.11	0.60	6,127	66,638
1906	2.30	4.32	91.42	98.04	1.43	99.47	0.40	0.13	0.53	4,743	66,382
1901	2.11	4.66	91.71	98.48	0.93	99.41	0.46	0.13	0.59	3,886	66,126
1896	1.66	3.96	93.24	98.86	0.66	99.52	0.37	0.11	0.48	3,788	63,924
1891	1.80	3.90	93.10	98.80	0.72	99.52	0.39	0.10	0.49	3,347	

TABLE B.0.2. Stage wise decomposition of enrollment in colonial India

TABLE B.0.3. Enrollment of students by grade level in 1917

	Total enrollment by grade level in year 1917 (in thousands)												
Province	A	B	1	2	3	4	5	6	7	8	9	10	Total
Madras	1.6	647.2	294.1	220.6	184.0	69.2	27.9	20.8	17.2	12.0	9.6	10.5	1,514.7
Bombay	184.0	77.9	125.3	108.2	90.4	68.3	33.1	20.1	15.2	0.3	0.2	—	723.2
Bengal	491.5	402.7	271.1	210.0	126.1	85.9	54.9	47.5	32.1	27.5	21.3	19.8	1,790.3
United Provinces	246.8	152.5	128.3	86.9	59.6	42.4	22.5	20.1	8.0	6.7	6.0	5.1	786.0
Punjab	—	—	160.7	67.8	55.3	41.8	35.7	16.4	12.5	9.3	6.0	4.8	410.4
Burma	121.3	—	112.6	67.2	38.2	21.8	9.9	6.8	5.2	1.1	0.9	0.7	385.6
Bihar and Orissa	352.5	152.5	90.4	68.9	22.9	18.4	10.1	9.2	6.0	5.0	4.2	5.0	745.2
Central Provinces, Berar	89.1	3.0	76.3	58.0	49.8	40.2	9.5	7.3	3.6	2.8	3.9	1.6	344.9
Assam	—	—	119.0	32.7	29.8	19.5	6.9	4.8	3.1	1.7	1.3	1.3	220.2
North-west Frontier	0.0	0.0	17.9	7.4	5.7	4.0	2.0	1.4	1.0	0.9	0.5	0.3	41.0
Minor administrations	7.7	2.3	8.7	5.7	4.9	3.9	2.4	1.8	1.5	1.0	0.8	0.7	41.5
India	1,494	1,439	1,404	934	666	415	215	156	105	68	54	50	7,003

TABLE B.0.4. Enrollment of students by caste and religion

		Share of enrollment in all institutions and levels by castes and religions (in percent)								
Province	Year	Europeans and Anglo Indians	Christians	Brahmins	Hindu non Brahmins	Muslims	Buddhists	Parsis	Others	Total student population (in thousands)
Madras	1917	0.52	6.95	11.14	69.72	11.08	0.03	0.01	0.55	1661.0
Bombay	1917	0.64	2.82	12.81	62.13	19.18	0.00	2.05	0.37	780.5
Bengal	1917	0.48	0.63	8.70	43.95	45.05	0.44	0.01	0.73	1918.4
United Provinces	1917	0.64	1.05	25.80	53.93	18.18	0.00	0.01	0.39	894.8
Punjab	1917	0.63	1.34	8.94	34.40	41.30	0.01	0.02	13.36	476.7
Burma	1917	0.88	4.08	0.08	0.98	4.20	89.70	0.01	0.06	592.5
Bihar and Orissa	1917	0.13	2.65	13.67	65.10	13.03	0.00	0.00	4.23	845.0
Central Provinces and Berar	1917	0.43	1.14	11.36	62.69	9.23	0.51	0.09	14.61	351.1
Assam	1917	0.09	5.26	7.91	50.41	23.77	0.04	0.00	12.48	233.9
North-west Frontier Province	1917	0.11	0.11	1.94	26.78	70.84	0.00	0.00	0.00	46.3
Minor administrations	1917	5.44	6.41	9.71	53.59	20.97	0.02	0.19	3.50	51.5
India	1917	0.54	2.95	11.53	51.92	23.24	6.91	0.22	2.70	7851.9
India	1912	0.51	2.75	12.10[1]	53.70[†]	22.88	6.04	0.26	1.80	6780.7
India	1907	0.57	3.14	12.10[‡]	53.70*	21.76	6.84	0.32	1.58	5388.6

† ‡ * Hindu disaggregated figures are approximate for the years 1912 and 1907 assuming a proportional split between Brahmins and non-Brahmin Hindus after accounting for the shares of all other caste and religious groups.

Hh. pc. cons. rank	Share of private educational investments in 1986 (in percent)					Change as of 2014 (about 30 years)				
	Bottom 50	Next 40%	Top 10%	Top 5%	Top 1%	Bottom 50	Next 40%	Top 10%	Top 5%	Top 1%
Andhra Pradesh	24.5	45.3	30.2	18.6	4.5	0.2	0.6	-0.8	-1.2	0.8
Bihar	33.6	45.9	20.4	13.8	3.6	-9.7	-6.1	15.8	11.4	4.1
Delhi	30.2	47.3	22.6	9.6	0.9	-7.8	1.3	6.5	9.5	4.2
Gujarat	30.1	40.2	29.7	15.3	2.6	-12.3	5.1	7.2	9.0	4.5
Jammu & Kashmir	39.1	39.8	21.1	12.4	2.4	-8.1	5.3	2.8	3.7	1.8
Karnataka	29.3	38.7	32.0	22.0	4.8	-8.7	5.3	3.4	-3.0	1.7
Kerala	32.7	42.6	24.7	10.1	4.0	-10.1	-0.7	10.8	12.8	-0.3
Madhya Pradesh	32.5	44.9	22.6	15.0	3.1	-15.6	-1.9	17.5	13.4	6.3
Maharashtra	27.3	41.5	31.2	19.0	4.3	-10.0	2.7	7.2	0.1	1.4
North-east states	39.4	42.6	18.0	10.0	2.8	-15.8	6.4	9.5	7.5	1.6
Orissa	34.3	34.2	31.6	17.4	4.1	-18.0	1.5	16.6	15.1	4.9
Punjab	30.2	49.7	20.1	13.1	2.3	-5.3	5.4	-0.1	3.0	2.1
Rajasthan	31.2	43.9	24.9	18.5	1.7	-6.0	-0.7	6.7	1.2	4.3
Tamil Nadu	22.2	44.3	33.6	16.0	3.7	2.7	3.4	-6.1	1.9	2.5
Uttar Pradesh	30.9	44.0	25.1	14.5	4.0	-11.1	-7.5	18.6	14.6	8.3
West Bengal	26.0	42.1	31.9	15.5	5.7	-5.2	-5.8	11.0	14.3	3.6
India	29.9	41.0	29.1	16.0	4.4	-11.6	3.9	7.7	10.4	0.2

TABLE B.0.5. Concentration of private educational investments in recent period

Hh. pc. cons. rank	Share of educational investments (public and private) in 1986 (in percent)					Change as of 2014 (about 30 years)				
	Bottom 50	Next 40%	Top 10%	Top 5%	Top 1%	Bottom 50	Next 40%	Top 10%	Top 5%	Top 1%
Andhra Pradesh	33.9	42.7	23.4	14.3	3.4	-0.9	0.7	0.2	-0.4	0.6
Bihar	35.3	45.3	19.4	13.1	3.2	3.5	-7.2	3.7	2.6	1.4
Delhi	32.8	46.0	21.2	9.1	0.8	-5.6	0.1	5.5	8.5	3.8
Gujarat	40.7	37.2	22.1	11.3	2.4	-9.8	3.3	6.5	7.6	3.3
Jammu & Kashmir	40.6	39.7	19.7	9.7	2.3	-4.6	4.8	-0.2	3.7	1.3
Karnataka	42.8	37.3	19.9	13.3	2.7	-13.8	2.3	11.4	4.9	6.2
Kerala	36.8	42.8	20.4	8.2	3.0	-8.2	-2.7	10.8	11.6	0.1
Madhya Pradesh	39.7	42.2	18.1	11.8	2.3	-7.3	-2.7	10.1	8.2	4.1
Maharashtra	31.1	41.5	27.4	16.5	3.6	-5.1	0.5	4.7	-0.8	1.0
North-east states	40.8	43.1	16.0	8.9	2.5	-6.2	2.0	4.1	3.9	0.5
Orissa	40.1	32.4	27.5	15.0	3.5	-6.1	2.5	3.6	5.6	2.1
Punjab	32.5	48.5	19.1	12.2	2.2	-2.7	3.5	-0.8	2.4	1.8
Rajasthan	35.8	40.7	23.6	18.0	1.5	0.5	-2.1	1.6	-2.1	3.2
Tamil Nadu	47.5	37.1	15.4	7.0	1.6	-15.9	7.8	8.1	8.4	3.6
Uttar Pradesh	32.6	43.2	24.3	13.6	3.7	-1.7	-8.0	9.7	8.8	5.7
West Bengal	30.5	41.8	27.7	13.5	5.2	2.8	-5.5	2.7	7.3	1.3
India	37.7	38.6	23.7	12.9	3.5	-8.6	3.5	5.1	7.7	2.9

TABLE B.0.6. Concentration of total educational investments in recent period