Sarah Nolan
Duke University
10.19.21

Roll Call: The Social Origins of India's Engineering Students

Abstract: Rapid economic change shapes the choices available to many of the world's young people. Yet it remains unclear which young people are gaining entry to new competitive opportunities, a question complicated by gaps in data, unfit measures, and the unknown influence of rapid change itself. This paper addresses these bigger questions by focusing on one elite context: engineering undergraduate degrees in India. I find five distinct subgroups within the student body, using Latent Class Analysis, a model-based technique for identifying hidden populations. A significant proportion of India's engineering students come from backgrounds of mixed advantage and disadvantage, with socio-economic disadvantages cooccurring alongside rural and/or low caste backgrounds. This complexity is less apparent using traditional one-dimensional measures, or only focusing on country level measures. Additionally, I find that the process of gaining access likely differs between institutional quality tiers. A higher proportion of groups eligible for affirmative action, and a lower proportion of women, regardless of socio-economic origins, attend top tier institutions. This divergent pattern suggests different attainment processes, and that enrollment policies may provide some narrow support for expanding opportunity. More broadly, these findings suggest that context specific, multidimensional approaches to social stratification in places experiencing significant change can both improve our understanding of status attainment and more directly inform opportunity enhancing policy.

1 Introduction

Young adulthood has always been a time of transition: individuals must make choices about education, work, and family. Yet in many places, seismic shifts in employment, education, migration and more mean that today's young people often face different choices than their parents faced at the same age, or even their older cousins and siblings. Some elements of this change are readily measured: new seats in college classrooms, new apartment buildings in cities for new workers in new jobs. Yet while opportunities are changing, it is harder to understand what this change means about the *distribution* of these opportunities. Today's growth is stacked on a foundation of long-standing inequalities. What are the chances a particular young person can access these new opportunities?

Measuring the distribution of opportunities proves far more complex than measuring their expansion. Estimates of social mobility summarize the extent to which the circumstances of individuals' early lives influence adult outcomes. Yet a large and growing literature finds that whether a particular society seems very mobile or immobile is highly sensitive to researcher measurement choices. Places experiencing rapid and novel social shifts in the very measures used to calculate social mobility– education, work, wealth–call for additional caution. Finally, a growing thread of research calls for greater attention to *how* social mobility occurs, in addition *to what extent.* While cross-country comparisons have long been dominant, elucidating changing processes will require more micro-level snapshots of these processes and the individuals who move through them. In times when the rules of the game are changing, what kinds of people are getting ahead, and how?

Answering these questions can help highlight in what ways desirable opportunities have become more diverse, where barriers remain, and tools for investigating contexts with similar dynamics. This paper focuses on one rapidly changing, data-constrained, and demographically consequential context – India's higher education system, specifically its undergraduate engineering institutions.

In this paper, I ask the following three questions:

1. What are the socio-economic origins of students attending India's engineering universities?
2. How do these socio-economic origins intersect with other traits known to be important to individual life chances in this context – gender, caste and religion, rurality?
3. Within the engineering student body, what factors distinguish students attending upper tier and lower tier institutions?

I investigate students' origins using an inductive method to describe multiple dimensions of an individual's background. Indeed, this analysis finds that social origins are not easily reduced to one continuum of low to high or one summary variable. More than half of the sampled students come from backgrounds of mixed advantage and disadvantage. I find examples of upward mobility at the margins: while most engineering students are from advantaged backgrounds, and most young people from disadvantaged backgrounds do not become engineering students, there remains a significant proportion of disadvantaged or mixed advantage young people within engineering schools. Women, Muslims, rural, and rural low-caste individuals remain underrepresented. These identities are inextricably linked with socio-economic origins. Engineering students mirror population level disparities, with a strong association between a student being rural and or low caste and being from more disadvantaged socio-economic origins. This finding provides support against the "creamy-layer" argument – that low-caste students in higher education are uniformly economically advantaged. In short, this holistic review of social origins suggests that disadvantaged students are not a few brilliant or lucky outliers, nor are they easily dismissed as a measurement anomaly.

However, comparing within tiers of engineering education reveals a different pattern of inequality. Women of all socio-economic backgrounds are particularly underrepresented in top-tier institutions. The opposite is true for students from low caste backgrounds, possibly reflecting the influence of affirmative action policies in funneling the small number of low caste students into better schools. For other groups (non-affirmative action caste groups, men, rural or urban students) socio-economic origins seem more strongly associated with top tier school admission.

This snapshot approach demonstrates that today's engineering students are a diverse group, a diversity that would be less apparent if taking a variable-oriented or country-level approach. The inductive approach taken here shows how disadvantages of different types come together to shape individuals' social mobility trajectories. What exactly the most salient disadvantages are and at what point they operate needs to be known more precisely, so appropriately targeted policies are put in place. Arising from a fluid underlying situation, the structure of disadvantage is also liable to change and shift, making regular examinations necessary.

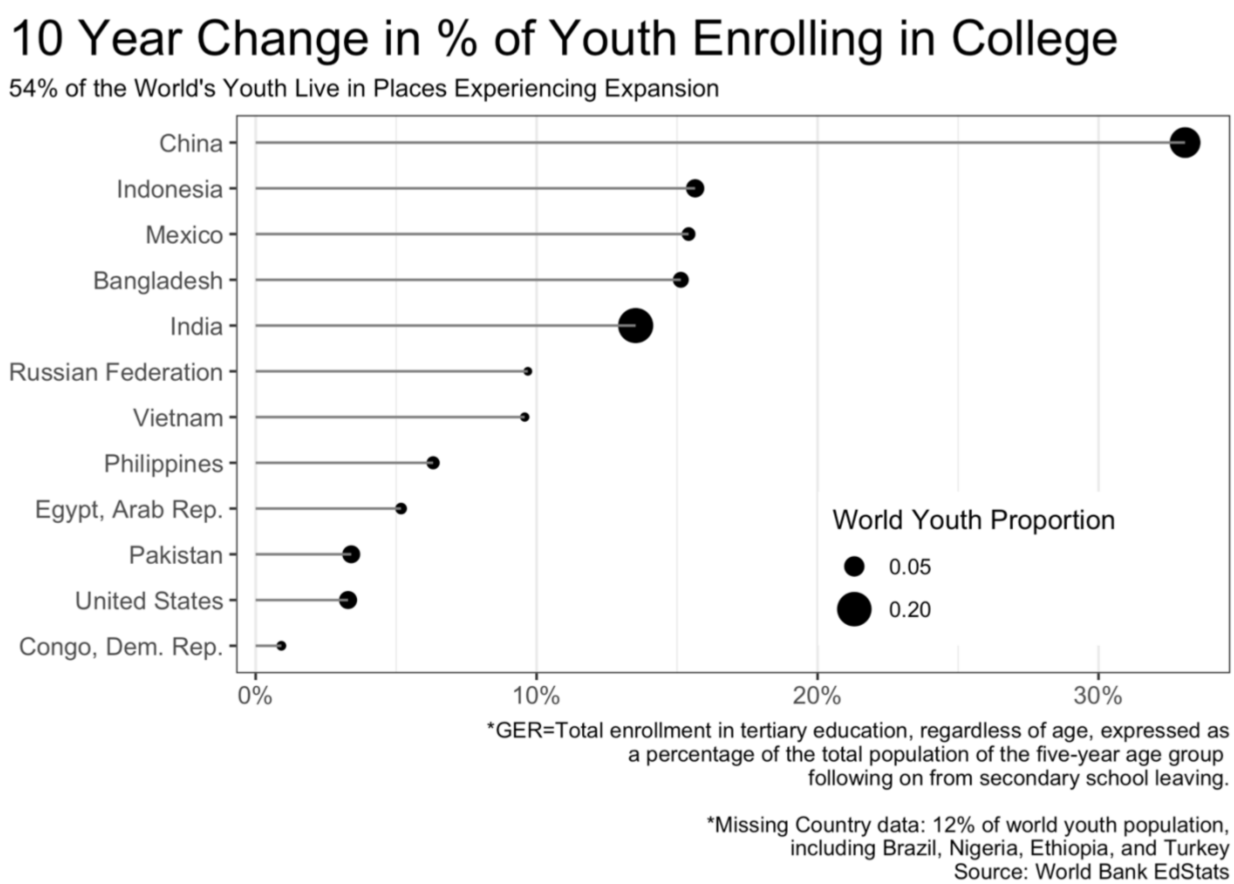
Despite persistent inequality across India's primary and secondary education systems, some disadvantaged students do gain entry to higher education, and institutions who have traditionally supported the most advantaged students must also consider how to support a more diverse student body. Additional space for policy supports include access to higher quality institutions, regulation of program quality, and job placement. Finally, this paper demonstrates that similar person-oriented, inductive explorations can help scholars and policy makers better understand social mobility in other pathways and contexts.

2 Literature Review

2.1 Educational Expansion and Equity

Most of the world's youth live in places experiencing tertiary educational expansion. In the past ten years a diverse set of places have seen their youth cohort's enrollment grow by over 10 percentage points. Figure 1 shows the change in the percentage of youth enrolled in tertiary education in places with over 1% of the world's youth cohort. This quick expansion both shifts the landscape of higher education, and the choices young people can make, with very little precedent from the preceding generation.

Figure 1



Rapid expansion in higher education enrollments can shape social mobility prospects for both individuals and society. A degree is increasingly a pre-requisite of professional, stable and high paying work (Montenegro and Patrinos 2014). Education is also associated with better health, family relationships and parenting (Oreopoulos and Salvanes 2011). Politicians, CEOs, and community leaders prioritize tertiary education as way to shape more productive, entrepreneurial, and politically content citizens. These society-level benefits are difficult to prove, yet they feature heavily in the mission statements of actors from UNESCO to grass roots nonprofits (Chabbott and Ramirez 2000, Hannum and Buchmann 2005).

Evidence suggests that while getting any education is good, getting higher education can be even better. The individual returns to higher education are quite high in many developing countries

(Montenegro and Patrinos 2014). This "college premium" appears to hold true even for young people from disadvantaged backgrounds, although evidence for this relationship mostly comes from the U.S and Europe. If an individual from a disadvantaged background attains a college degree, their adult outcomes become far more equal to their advantaged peers, across a variety of indicators (Hout 1984, Hout 1988, Breen and Jonsson 2007, Torche 2011).

Yet educational expansion alone does not bring more equal outcomes for all young people. Instead, growth and equity concerns are "two sides of the same coin" (Walters 2000). Even when the educational system expands, it does not necessarily change the criteria used to evaluate students, nor the support available to young people hoping to meet that criteria. Young people can still leverage their family advantages to get ahead of their peers with fewer resources.

Country level data make clear that many young people live in places experiencing tertiary expansion. Yet understanding if this has meant more equal enrollment among underrepresented groups is harder to track. Prior empirical explorations of educational growth and equity have found varying results, offering few strong assumptions about trends today (Shavit and Blossfeld 1993, Hannum and Buchmann 2005, Shavit 2007, Bloome, Dyer et al. 2018). Secondly, we might want to know why or how more equitable attainment occurs, and how perhaps it could be supported by policy. Addressing upward mobility requires not only documenting how many students matriculate, but also where in the social hierarchy they come from – and through what channels and processes?

2.2 Measurement Challenges in Social Status and Mobility

Describing social hierarchies and placing individuals within them is one of the most active debates within the social sciences (Grusky 2001). Some studies measure social status through one variable, for example, an individual's income or occupation. Others create more complicated syntheses –socio-economic status indexes, occupational prestige scores, or class schemes. Still others, faced with insufficient data, use one variable to estimate another, for example using education to infer an individual's income. Torche provides a useful review across disciplinary histories (Torche 2015).

Scholars have struggled to translate these approaches to rapidly developing contexts. A lack of longitudinal, population representative datasets challenges even basic descriptions of any variable of interest. This lack is more acute for scholars studying intergenerational mobility. Only a few surveys ask respondents about their non-resident family and origins, leading to long-running debate on "co-residency bias" in studies which only measure mobility between co-resident parent -adult child pairs (Emran, Greene et al. 2017, Narayan, Van der Weide et al. 2018).

Even beyond these data constraints, patterns of social and economic life also differ in ways that challenge estimating static, individual social status. Work and assets are marked by dynamism over time within families: places with informal work often earn a living through a variety of efforts, including farming, labor, and small-scale vending, rather than having a few salaried wage earners. This makes for fluctuation in income with the seasons, and more volatility over years (Chambers 1995, Krishna 2010, Rains and Krishna 2020). Furthermore, this economic cooperation within extended households makes separately assigning individual income and occupation challenging (Emran and Shilpi 2019). Social class schemes designed in the Global North may poorly fit these places for similar reasons. For example, landownership does not necessarily signal economic security in developing countries, where many farmers are working small subsidence plots. Business ownership can similarly signal anything from a temporary food cart to a formal consulting practice (Torche 2014).

A "coarse" distribution provides a third challenge to measuring status. In many developing countries, wide-ranging inequality means many individuals share a common status. For example, in India, a large proportion of adults share a few characteristics: basic educational attainment, agricultural work, low number and variety of assets, with only a small elite segment with more advantage attributes -see for example (Iversen, Krishna et al. 2017, Asher, Novosad et al. 2020) and a review (Iversen, Krishna et al. 2019). This "coarseness" can mean assigning the same position to half the population, even if the lived realities of individuals within these coarse categories are quite diverse.

Scholars focused on developed countries have had parallel debates. One strand of the literature questions whether measures, particularly social class schemes, currently in use reflect today's relevant inequalities (Weeden and Grusky 2005, Grusky and Weeden 2008, Weeden and Grusky 2012, Savage, Devine et al. 2013). A related thread examines multiple dimensions of status within individuals, questioning the extent to which dimensions correlate and diverge across time and place (Bukodi and Goldthorpe 2013, Jonathan, David et al. 2016, Blossfeld 2019).

While the study of intergenerational social mobility is in part measuring population level associations between parents and adult children, it is also an investigation of "social origins", "the conditions and circumstances of early life" and the ways in which they "constrain success in adulthood" (Hout 2015). This view is inherently inductive, as the relevant constraints of social origins "vary by time, place, and subpopulation." Conceptually, social origin conditions can include an individual's family, neighborhood, local policies and much more.

This inherent complexity in origins poses measurement challenges. Social class schemes offer one approach. However, these schemes are deductive, applied by the researcher to the data at hand. These schemes may fail to accurately map onto to developing and changing contexts. Crafting indexes based on correlations between variables requires a wide range of numeric data, for example annual salary figures, wealth, or years of education. In places with limited data and substantial coarseness within distributions, this approach may not prove useful.

Furthermore, the question of how different young people get ahead or get stuck is distinct from answering where upward mobility is highest, or the role one variable plays in inequality. It's a question we can approach through more micro-examinations of specific pathways. This paper developed an inductive way to classify different types of origins, well-fitted to the specific population: engineering students in India in 2019. This approach can both highlight possible barriers to upward mobility and inform how we can bring more relevant measures to national level comparisons.

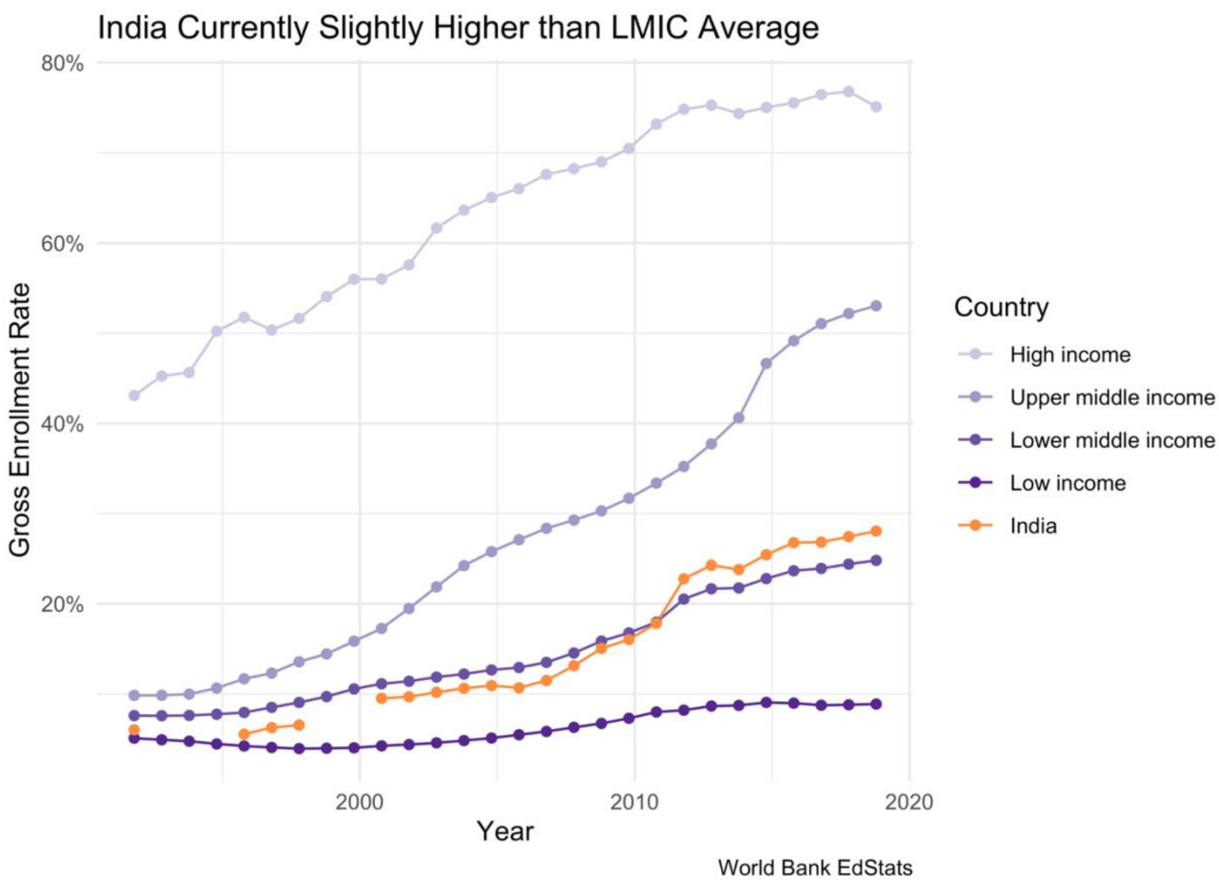
3 Context

3.1 Higher Education and Upward Mobility in India

Engineering education in India presents a useful case to explore educational attainment in times of change. India's expanding education, widening inequality and apparent low upward mobility makes educational access a pressing concern. These trends are also emblematic of other contexts experiencing rapid and unequal growth.

India has one of the largest higher education systems in the world. Students in India make up about half of all tertiary students in lower middle-income nations, and about fifteen percent of all tertiary enrollments globally (Word Bank Ed Stats). India's higher education systems is not only large, but rapidly growing. Higher education Gross Enrollment Ratios (GER) capture the proportion of "college-age" (18-23 years old) individuals who are enrolled in higher education. In 2019, India's

GER is 27%, increasing from just 9.5% twenty years ago (WorldBank 1950-2050), bringing it slightly higher than the average GER for lower middle income countries.
Figure 2



Yet higher education is only one of many big shifts in India's recent history. A thread of expansion, and expanding inequality, runs through all available economic indicators. GDP per capita rose through much of the 2000s, and poverty fell steadily during the same period (Bank 2019). But while average living conditions have improved, inequality in consumption and income appears to have risen along with GDP per capita (Dang and Lanjouw 2018). Wages have been rising at the top, yet India's much hyped tech, finance, real-estate and other well-paying sectors only employ a small number of workers (Dang and Lanjouw 2018). One analysis claims the top 0.1% of earners captured a higher share of total growth than the bottom 50% (Chancel and Piketty 2017). The lowest paid 50% of workers only received 12% of all wages paid (International Labour Office 2016). Almost 90% of working individuals are in informal positions (NCEUS 2009). Employment in and income from agriculture has been slowly falling, yet it remains a common economic activity in India (OECD and Relations 2018). These changes complicate measures of social status.

India has some of the best higher education institutions in the world, notably the longstanding public "IIT" system. Yet existing evidence suggests that most young people with less educated parents do not gain higher education, limiting its potential as a driver of upward mobility. One thread of the literature considers relative intergenerational educational mobility: the relationship between parent's educational attainment rank within their cohort and their child's educational attainment rank within their own cohort. This relative approach captures mobility patterns distinct from growth in education between generations. When parent-child relative ranks are highly related, it suggests low educational mobility, and therefore that the children of less educated parents achieve less education than their peers whose parents had higher attainment. This is a unidimensional, variable-oriented approach- measuring the association of parent's education on their children.

Similar to the literature on such associations across the developing world, the conclusion about intergenerational educational mobility in India is unclear, leaning toward low upward mobility. Truncated co-resident samples of parent-adult child pairs can lead to biased estimates, and several papers have noted that correlation and regression coefficients for the same set of individuals can provide contrasting estimates (Emran, Greene et al. 2017). A few earlier studies found increasing mobility over time (Jalan and Murgai 2008, Hnatkovska, Lahiri et al. 2013). However, most recent evidence points to stagnant rates of mobility, or India lagging behind similar economies. (Emran and Shilpi 2015, Narayan, Van der Weide et al. 2018, Asher, Novosad et al. 2020). Urban environments appear to have higher upward mobility than rural, a difference which is particularly stark for women (Emran and Shilpi 2015, Vaid 2016, Asher, Novosad et al. 2020). Scheduled Caste and Tribe groups seem to have become more upwardly mobile over time, possible related to affirmative action (Hnatkovska, Lahiri et al. 2013, Asher, Novosad et al. 2020). Meanwhile new evidence suggest that Muslim Indians have experienced correspondingly lower levels of mobility (Asher, Novosad et al. 2020).

Another approach examines the proportion of young people from various backgrounds that do attain higher education, a series of estimates known as transition matrices. Asher et al provide one recent set of estimates from the India Human Development Survey in 2012. For sons born in 1980-1989, 17% achieved any higher education, yet this rate varies significantly by father's education. Only 5-8% of sons with fathers with less than primary education (about half of all fathers) get any higher education, while that number rises to 65% of sons with fathers who themselves have higher education(Asher, Novosad et al. 2020). Taken together, these population estimates suggest a great deal of intergenerational stickiness and little upward mobility.

3.2 Describing Engineering Pathways

Given the fast-changing nature of education and work in India, our knowledge of the composition of each node on the pathway remains imperfect. The conditions of early life, primary and secondary schooling are highly unequal in India, with rural versus urban, private vs government schools, regional differences, and an individual's caste, religion, and gender proving to be important cleavages.(Desai and Kulkarni 2008, Vaid 2012)

The pathway to engineering education narrows during the final two years of secondary school, when students must score well on lower secondary exams and select "science" as their focus. Students must score above 50% on their class 12 exams to graduate. At this point, many seek out "coaching" in private institutes to prepare for entrance examinations. This process is intense, and students often move across the country to a coaching hub. Coaching is poorly regulated, time and money intensive, and deeply competitive (Ørberg 2018). While the flagship national schools (IITs and NITs) share a two-step entrance examination, students can alternatively elect to take state-based entrance examinations, or an exam for a particular network of government aided or private institutions, which then make them eligible for different seat allotment schemes.

Admission is based primarily off a student's rank on the relevant entrance exams. This rank determines the order students can select from remaining seats in participating institutions. Top scorers get early priority. Quotas, called "reservations", make some seats eligible to students based on caste, gender, and/or state residency. Therefore, students from these reserved groups can and do gain admittance to more competitive institutions, despite lower entrance exam scores (Bagde, Epple et al. 2016) . The "management" quota system serves a different set of prospective students: those who negotiate with private institutions to pay a much higher fee and gain admittance without an entrance test score, another avenue for the wealthy to seek higher quality institutions.

Secondly, reservation admits are generally from poorer families than their higher caste peers, despite recurrent political rhetoric about reservations "cream-skimming" students from rich, lower-caste families (Bertrand, Hanna et al. 2010). Research conflicts on how reservation admits experience college with Badge, Epple et al finding they perform equally well in exams and graduation rates, while Betrand, Hanna et al and Frisancho and Krishna finding lower classroom performance, job placement outcomes, and heightened mental health problems (Bertrand, Hanna et al. 2010, Bagde, Epple et al. 2016, Frisancho and Krishna 2016).

Yet as Frisancho and Krishna write is its perhaps "the interaction of poverty and SC/ST status that is most harmful" to grades and future earnings of minority students. They find that economically well-off lower caste admits, although few in number, have college outcomes commensurate to their higher caste classmates. Krishna in a survey of 671 students at five engineering colleges finds underrepresentation of low caste, female, rural, children of less educated parents, of agricultural workers, and of low asset homes. Students with any combination of these disadvantages are hardly present. Through interviews, the few "outliers" with multiple forms of disadvantage described their difficulties in navigating the college preparation process while also weathering family hardship (Krishna 2013).

Finally, a few studies have examined the latter stage of the engineering pathway, the composition of employed engineers. This work suggests that under-representation of those with any disadvantage persists into employment. Working engineers tend to come from elite families, with parents who worked as salaried government employees, engineers, and those with inherited family wealth (Krishna and Brihmadesam 2006, Fuller and Narasimhan 2007).

Navigating this process takes both tenacity and intense self-advocacy. Regulatory agencies have not been able to keep up with the rapid pace of expansion, and many unregulated institutions are more than happy to take students' money and time in exchange for low quality education, not to mention the even less regulated test-prep industry (Upadhya 2016, Ravi, Gupta et al. 2019). For students whose family members may have only a primary school education, navigating to a quality institution can be a bewildering and financially ruinous experience. Indeed, recent qualitative evidence suggest that savvier, often wealthier and better connected, families navigate to higher quality institutions, while lower quality schools happily admit less privileged students alongside their government funded scholarships(Upadhya 2016).

Taken together, the relationship between growth and upward mobility in higher education remains unclear in India. Many of the studies listed here focus on college cohorts prior to 2010-2020 expansion. An up-to-date description of the types of students attending, considering the intersections of socio-economic status, caste, religion, gender and location of origin, could provide important insight on both social mobility and how policy makers may be able to improve it.

4 Data and Analytic Strategy

4.1 Data

Data on current engineering students comes from a novel survey I distributed in the summer of 2019. The survey was distributed to students via the Aspiring Minds Computer Adaptive Test (AMCAT). Students elect to take this test to demonstrate their proficiency to employers in IT, communications, engineering and more. Institutions sponsor exam days for their students, so students do not pay an exam fee. Students were given the opportunity to participate in this survey after completing the AMCAT test, on the same computer. AMCAT tests and the appended survey took place on institution-specific test days.

While my sample was not drawn at random from all institutions nor all students within institutions, the 32 institutions represent a range of public and private institutions and geographic locations. In comparison to the overall universe of engineering institutions, the universities in this sample skew toward the medium to higher quality tier, based on rankings described in the following section. At the time of writing, there was no available representative dataset on India's engineering students. Given this deficit, I argue that this sample provides a useful sample on engaged students at the broad middle range of engineering institutions.

About 16,000 students opened the survey, of which 10,000 meaningfully engaged with the survey. While I do not know characteristics of the unknown number of students who did not open the survey nor those 5,000 who opened the survey and did not answer the first question, patterns of non-response within the survey suggest high engagement among all students upon starting the survey. Nonresponse rates for particular questions are low, between 1-4.% I find similar patterns in responses to socio-economic questions, a module in the middle of the survey, among students who did and did not complete the final section, on demographics. For more details on the sample and non-response rates, please see the **appendix Z.**

I restricted my analytical sample to students pursuing a Bachelors in Engineering (B.E) or Technology (B.Tech) who answered questions about their families and earlier life, demographics, and who I could link with a university. This reduced my sample size to about 8,400 students. I remove a further 300 students who provide difficult to categorize responses[1] to the SES questions, leaving my analytical sample at 8,109.

To assess my sample against the engineering student body more broadly, I use available proportions from the All India Council for Technical Education (AICTE) from the 2018-2019 academic year. I find my sample has a higher proportion of women. I also find a slightly higher rate of individuals from "minority" religions, a government designation that includes many Non-Hindu religions. Examining caste, my sample has similar levels of scheduled tribe and caste students, and a slight lower proportion of OBC caste members. Given the lack of micro-data on engineering students in India, I proceed, noting these slight differences.

[1] These hard to categorize responses include reporting a parent who was disabled, retired, or deceased, during the respondent's childhood, as well as any respondent who could not remember the answer for any of the five variables included in the LCA. In essence, I lose 300 respondents from among those who did try to answer by using list wise deletion.

Table 1

Trait		AICTE	This Sample
Gender	Female	29%	37%
Religion	"Minority Religion"	7%	12%
	Muslim	Not Provided	3%
Caste	Scheduled Tribe/ Schedule Caste	15%	13%
	OBC	35%	28%

Analyses on the socio-economic origins of students and gender, caste and rurality compare my sample with a general age cohort from the Indian Human Development Survey (IHDS II), a nationally representative household survey. Within the IHDS, I select a similar age cohort of all Indians, born in 1996 - 2001. IHDS II collected data in 2012, meaning the cohort of interest was about 11-16 years, old, roughly the same age I ask my respondents to recall. This cohort includes an unweighted 26,000 individual young people within sampled households, representing a weighted 174,000,000 Indian young people. Table 3 below describes my sample of engineering students and the IHDS sample. This quick snapshot shows a student body quite different in composition to the general age cohort, more male, urban, and General Caste.

Table 2

Characteristic	Engineering Sample N = 8109[1]	IHDS2 N = 26155[1]
Gender		
Female	37%	48%
Male	63%	52%
Caste or Religion		
Muslim	2.8%	15%
Other	9.0%	<0.1%
ST/SC	12%	30%
OBC	28%	36%
GC	46%	17%
CSJB	2.8%	1.4%
Location in Early Teens		
Rural	34%	71%
Urban	66%	29%

[1] Weighted (%), Unweighted N [2]

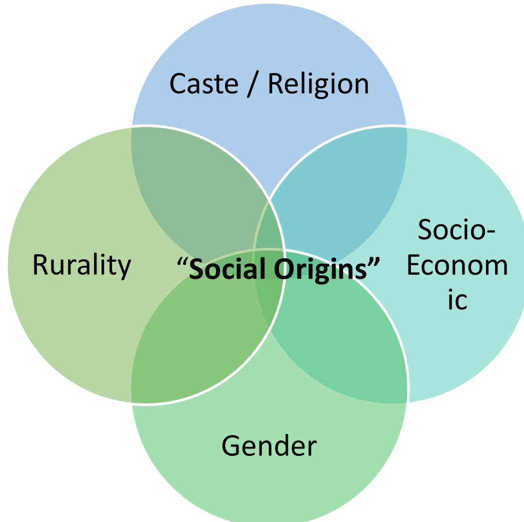
4.2 Analytical Plan

Considering the complexity of a situation where disadvantage can arise from multiple directions, I consider an individual's "social origins" in four interrelated dimensions. This paper's contribution focuses on defining socio-economic origins (SEO), as these are the attributes whose meaning for relative social status are most in flux, as the education, occupation, and assets of today's youth and their parents have changed significantly from previous parent-child cohorts. While the relative meaning of caste, religion, geography and gender are also in flux, these attributes are more consistently measured in administrative and scholarly work, and their changing meaning is partially driven by socio-economic resources. In order to understand these changing interactions, I focus first on defining of SEO in the first two analyses, and then intersecting this identity with the remaining three origin traits in the second two analyses.

Table 3 "Social Origin" Components

[2] Note that caste and religion proportions for my sample are slightly different between AICTE and IHDS comparisons. The two datasets follow different caste and religion definitions, which I match in each comparison. For the remainder of this paper, I follow the IHDS caste and religion combined social group variable I follow the IHDS coding practices for "groups" which sorts certain individuals by religion or caste depending on the combination of these two variables. CSJB refers to Christian, Sikhs, Jains and Buddhists, who are outside the caste system and are the most common non-Muslim religious groups. "Other" is my sample is mostly students who self-defined their religion as "India" or "humanity" while not providing a caste group. – for more information see https://ihds.umd.edu/social-groups

Figure 3 "Social Origin" Components



4.2.1 Variable-View of Socio-Economic Origins

First, I offer a descriptive look of the socio-economic origins of students in comparison to the general age cohort, using five variables describing the respondent's parent's education, work, and assets.

4.2.2 Latent Class Analysis of Socio-Economic Social Origins

I next use latent class analysis (LCA) to empirically identify student's socio-economic origins. LCA offers a statistical way to identify unobserved subgroups (latent classes) within a population based on observed information. In simple terms, LCA offers a way to summarize a large contingency table, listing the frequency of observed combinations of relevant categorical variables by individual. Instead of isolating the relationship between particular "independent" variables and outcomes of interest, LCA instead empirically describes types of individuals present in my sample. LCA can help us understand how observed measures come together to form complex constructs (e.g. "health", "risk", or here "social origins").

Using LCA allows socio-economic origins to become more nuanced than simply one continuum of disadvantaged-advantaged, and offers more insight on the complexities of individual experience. As such, methodologists have described LCA as a "person oriented" approach, in contrast to more commonly used "variable oriented" approaches to modeling relationships (Laursen and Hoff 2006, Collins and Lanza 2009). This paper does not decompose the role of each variable in explaining educational attainment, nor does it seek to assign one number to describe mobility rates in a population. It instead develops a typology of individuals, highlighting *patterns of socio-economic social origins*. The remainder of this paper refers to latent "classes" as latent "subgroups" to avoid confusion with the concept of "social class" in sociology

LCA is inductive, in that the model identifies patterns present in individual responses within this sample, not based on prior theory, a useful tool for a setting experiencing rapid change While more traditional social class analysis would offer a similar multi-dimensional approach, there is no consistent class scheme for much of the rapidly changing world.

The number of subgroups the model estimates, as well the names and interpretation of these subgroups, is researcher derived. LCA is probabilistic, and estimates are made with some uncertainty as subgroup membership can't be directly observed. I document the iterative process model building process and uncertainty estimates in both the analytical section and appendix.

4.2.3 Intersections with SES Social Origin Subgroups and Identity Traits

A third analysis builds off the LCA derived subgroups, and explores how gender, caste and religion, and rurality intersect to help better describe the engineering student body. I describe both how these traits differ among SES origin subgroups, and also how these traits differ among estimated members of each subgroup. As subgroup membership is a latent typology, I use a model-based approach to generate estimates of gender, caste, and rurality within subgroups, incorporating the uncertainty present in my base LCA model. To examine the inverse relationship, how different traits are associated with subgroup membership, I use multinomial logit, with subgroups as the dependent variable and gender, caste and rurality as the independent variables.

4.2.4 Associations Between Social Origins, Identity Traits, and Institutional Quality Tier

In a final analysis, I use latent subgroup membership to estimate probability of attending "top tier" institutions for different types of engineering students – examining combinations of socio-economic origins and rurality, caste, and gender.

5 Analysis

5.1 Descriptive Analysis of Socio-Economic Variables

Table 4 describes the five socio-economic origin variables in this analysis. Variables have been broken into three or four categories, reflecting key distinct levels within the engineering student population. The level order from Low-High is researcher imposed and color-coded for consistency. These categorizations elide over significant diversity within broad categories: farmers and small business owners can have varying degrees of size and success. Salaried work can range from low-level government clerks to well compensated software engineers. What unites this broad group is a steady paycheck, a meaningful distinction in India's economy where 90% of the workforce may be informal (NCEUS 2009). However broad, these categorizations offer a useful starting point, and alternative divisions do not alter the direction of the results. For more information on variable definitions, please see the appendix.

Table 4 Socio-Economic Variable Level Definitions

<table>
<tr><th>Level</th><th>High</th><th>Med-High</th><th>Med-Low</th><th>Low</th></tr>
<tr><td>Father's Education</td><td rowspan="2">Master's Degree or PhD</td><td rowspan="2">Bachelor's Degree</td><td rowspan="2">Any Secondary-Diploma (short post-secondary)</td><td rowspan="2">Middle School or Less</td></tr>
<tr><td>Mother's Education</td></tr>
<tr><td>Father's Occupation</td><td>Salaried work – educators, government, engineers etc.</td><td>Self-Employed Business Owner</td><td>Farmer</td><td>Daily wage labor</td></tr>
<tr><td>Mother's Occupation</td><td colspan="2">Salaried Work</td><td>Homemaker</td><td>Farmer or daily wage labor</td></tr>
<tr><td>Assets</td><td colspan="2">Any upper-tier asset (computer, car etc.)</td><td>Mid-Tier Assets (TV, piped indoor water)</td><td>No upper or mid-tier assets</td></tr>
</table>

Across each variable, the engineering student body looks quite different than the general population. In line with the literature on India's social stratification in general, and engineering students in particular, the distribution of individuals within categories is quite "coarse" -the engineering cohort predominately falls in the top one or two categories, while the general age cohort falls in the bottom coded category.

Focusing on the bottom coded level, engineering students and the population are most similar in terms of mother's occupation: an engineering student is five times less likely to have a mother working as a farmer or daily wage laborer than the general population. The two groups diverge to the greatest extent in father's education – an engineering student is eight times less likely to have a father with a middle school education. Comparing at top coded levels, engineering students are a striking 19 times more likely to have mothers with a master's or PhD than the general population, and 11 times more likely to having fathers with the same.

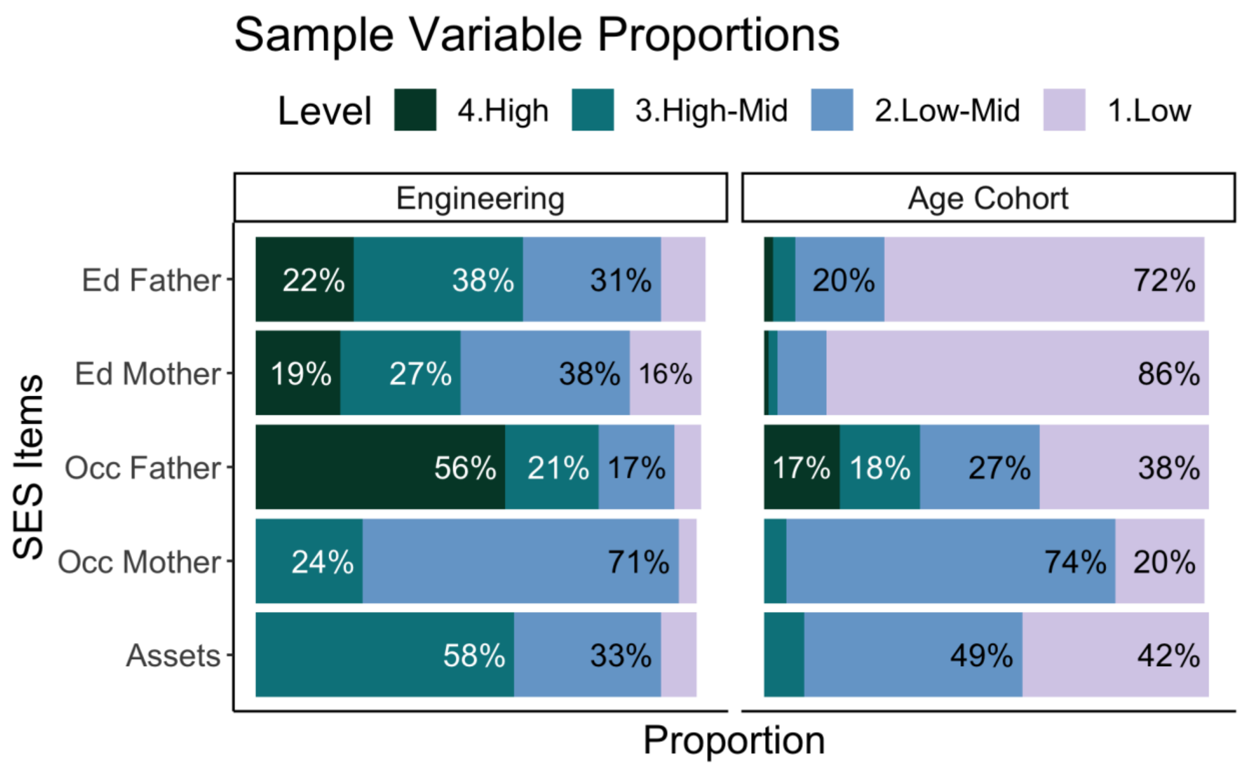
Variable by variable, engineering students appear to come from quite privileged backgrounds. Yet this approach leaves unresolved how these variables come together to form an individual's social origins. If social origins are the concept of interest, the "conditions and circumstances of early life", then we must examine not just the frequency of variables in a population, but the frequency of types of individuals in a population. For this, we move to the next analysis

Table 5 Variable Descriptive Statistics

Characteristic	Engineering Sample N = 8109[1]	IHDS2 N = 26155[1]	p-value[2]
1.Father's Ed			<0.001
Middle School or Less	9.8%	72%	
Any Secondary-Diploma	31%	20%	
Bachelor's	38%	5.4%	
Master's or Ph.D	22%	2.3%	
2.Mother's Ed			<0.001
Middle School or Less	16%	86%	
Any Secondary-Diploma	38%	11%	
Bachelor's	27%	2.2%	
Master's or Ph.D	19%	0.9%	
3.Father's Occ			<0.001
DWL	5.5%	38%	
Farmer	17%	27%	
SEBO	21%	18%	
Salaried	56%	17%	
4.Mother's Occ			<0.001
DWL or Farmer	4.2%	21%	
Homemaker	71%	74%	
Salaried	24%	5.5%	
5.Assets			<0.001
Bottom Tier Assets Only	8.2%	42%	
Some Mid-Tier Assets	33%	49%	
Top Tier Assets	58%	9.0%	

[1] Weighted (%), Unweighted N
[2] chi-squared test with Rao & Scott's second-order correction

Figure 4



5.2 Socio-Economic Origins Typology for Engineering Students

Using Latent Class Analysis, I identify types of socio-economic origins based on observed response patterns to the five observed socio-economic questions. Engineering students faced with five questions, two with three response levels and three with four response levels could respond with 546 possible combinations of responses (4*4*4*3*3). In fact, 44433 (all highest) and 44423(all high except homemaker mother) only comprise 7% of all responses, and 11111 and 22222 each only make up less than 1%. A quick glance at these response patterns suggests complexity, but it's difficult to make sense of 546 possible combinations, or even the 329/546 patterns I observe in this sample. LCA analysis helps make sense of these possible response patterns, by using these combinations to suggest subgroup membership, a latent variable underlying the observed five variables. Instead of 329 types of individuals, LCA identifies a much smaller number of subgroups.

Each individual in the population is assumed to be a member of one and only one latent subgroup, which are mutually exclusive and exhaustive (Collins and Lanza 2009). Yet we cannot directly observe an individual's latent subgroup membership, so we instead assign each individual probabilities of belonging to each of the subgroups identified, based on their response pattern. The researcher specifies the number of subgroups, a process I detail in appendix X. For the engineering sample, five subgroups represented a good balance between model fit and parsimony.

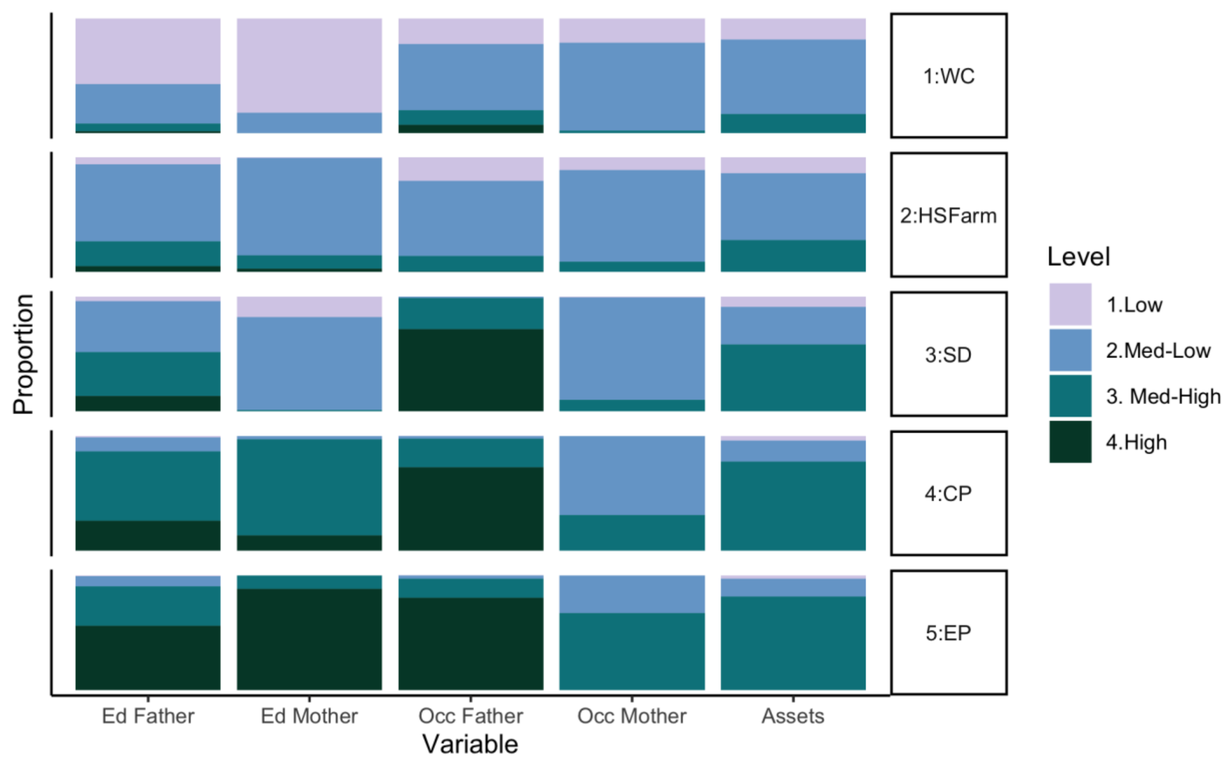
We can describe the identified social origin groups in terms of their variables using item response probabilities. The probability of response r_j to observed variable *j* conditional on membership in latent social origin group *c*, or $\rho_{j,r_j}|c$ is an item response probability. For any given variable *j* conditional on social origin group *c*, the probabilities for each response r_j sum to one. The item response probabilities for my base model are displayed in Figure 5 below.

Each row is a subgroup, and each column a variable. The multi-shaded bars within each row show item response probabilities: the conditional probabilities of a respondent falling into each response level for that variable, given that the respondent belongs to that latent subgroup. For example, the top leftmost cell shows that individuals in the "1: Working Class" have an estimated probability of 19%, 62%, and 18% for being in low, mid, and high asset levels respectively, summing to a probability of 1.

More formally $\rho_{ed\ father, low\,|\,1:WC} = .57,\ \ \rho_{ed\ father, medlow\,|\,1:WC} = .34,$
$\rho_{ed\ father, medhigh\,|\,1:WC} = .06,\ \ \rho_{ed\ father, high\,|\,1:WC} = .02$

To see the estimated probability for the other four variables for "Working Class", read the first row left to right. To see instead the estimated probabilities for each father's education level for each subgroup, read down the first column.

Figure 5 Item Response Probabilities



While the LCA model estimates these item response probabilities within subgroup, the subgroup names and ordering are researcher-assigned. I roughly order the groups 1-5 by probability of having a father with the highest education level – master's or PhD.

The first subgroup I call **1.Working Class (1:WC).** Students in this subgroup have a low probability of having had any top tier assets, both parents likely only gained a middle school, or occasionally, high school education, and fathers mostly work as laborers and farmers. Mothers almost uniformly work at home or as a farmer or laborer.

The second group I call **2. High School Educated Farmers (2:HSFarm).** This group has a similar profile to the first, except parents are very likely to have at least a secondary school education, and work as laborers, farmers, or a occasionally running a business.

Subgroup **3. Strong Dads (3:SD)** members have a higher probability of top tier assets than the previous two groups. Fathers are more likely to have a college degree than the previous two groups, and either run their own business or are salaried workers. Mothers, however, are more likely to have a high school education or lower, and are mostly homemakers. This mismatch in education and work between parents is present in all groups, but it is most marked in this group.

In subgroup **4. College Educated Parents (4: CP**), it's very likely the student had a top tier asset in the home growing up. Both mothers and fathers likely have a bachelor's degree, and fathers work, like the previous subgroup, is mostly salaried or running a business. Mothers in this group also may work for a salary outside the home.

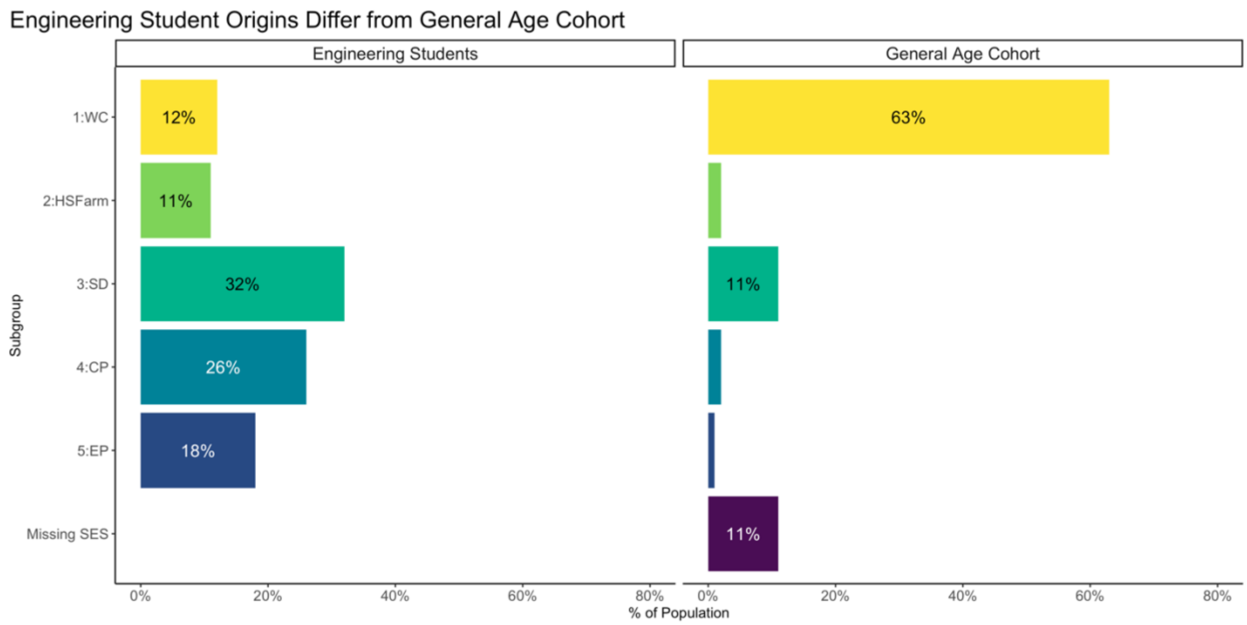
Finally, subgroup **5. Elite Parents (5:EP)** is distinguished from the previous group by the high probability that parents have a postsecondary degree, and that mothers in this group are the most likely to work as salaried position.

It's worth noting the diversity within these subgroup estimates. For example, the first group, despite being marked by low education levels and high probability of father's working as farmers or laborers, does have a thin sliver of college degree and salaried work. It's possible a few individuals with a college degree are working as farmers in this group, or middle school educated individuals have gained a salaried job. Some variable levels appear in multiple subgroups -for example, father's working as self-employed business owners appear in each subgroup, possible reflecting the heterogeneity of self-employed individuals. Despite this variation within group, the groups are still clearly unique from one another.

A second set of parameters estimates the population proportion of each social origin subgroup named above. These proportions are displayed below, in figure X. These proportions underscore that social origins are more than a binary of poor and wealthy. Each subgroup finds representation at over 10%.

Using the five variables broken into the same three or four categories and the same item response probabilities, I estimate the subgroup prevalence for this general age cohort using IHDS II data. While the population proportions for the engineering sample are model derived and estimated with uncertainty, I estimate the general age cohort proportions by assigning individuals to a subgroup based on their highest posterior probability ("modal assignment"). However, by "rounding up" to a probability of 1, I no longer account for the uncertainty inherent in assigning a latent subgroup. This can produce attenuated estimates (Bray, Lanza et al. 2015). Nevertheless, this approach allows me to directly compare the subgroup structure that best describes the engineering student body to the general age cohort. An analysis more interested in the precise subgroup structure of the general age cohort would be better executed by creating a distinct LCA.

Figure 6 Socio-economic Subgroup Population Proportions



While only suggestive, this comparison shows that most individuals in this age cohort are from "working class" origins. Having a father with a high school education and a salaried job, would distinguish your origins from most in the general population, yet this origin is unremarkable in the engineering student body. The across-the-board privilege of the 5. Elite Parents or 4. College Educated Parents is vanishingly rare across the country, but quite common in engineering classrooms. This takeaway is consistent with that of the variable oriented approach, and prior work on higher education and engineering in India.

Nevertheless, these social origin subgroups illuminate important distinctions within the student body. Subgroups 4: CEP and 5: EP are both quite advantaged, but only those students in 5:EP are likely to have grown up with parents with post grad degrees and to have a mother working in a salaried position. Students in 3: SD are similarly privileged to have high assets and a father in salaried in work, but their mothers are likely quite different, having only a high school degree. While small, a fifth of the students belong to groups 1:WC and 2:HSFarm, whose background certainly diverges from those of their peers. It's quite possible all of these differences between groups led to quite different experiences navigating toward engineering college. The following analyses delve deeper into the different demographics and experiences within each group.

5.3 Intersecting Identities

Gender, caste, religion, and rurality of origin are all consequential to young people people's trajectories. This section uses two distinct approaches to explore the intersections between SEO subgroups and these traits. How do the SES subgroups among engineers differ in their gender, caste, and rurality composition? How do these students with a given trait differ in their SEOs?

In this paper, I examine caste as operationalized by various government schemes, including higher education affirmative action: general caste (GC), other backward castes (OBC) and Scheduled Castes and Scheduled Tribes (ST/SC)/ The broad role and experience of caste in Indian society is somewhat locally determined and dynamic.[3] Very briefly, caste groups have shaped family, work and

[3] See for example Jodhka 2018, or Vaid 2014

business, housing, and political voice in India. General Caste (GC) individuals have occupied more privileged places in society, while Scheduled Castes (SC) and Tribes (ST) have traditionally had the lowest status positions. Other Backward Castes (OBC) form a wide-ranging middle of the hierarchy. These caste categories are government determined, and the site of political activism, as they determine government caste quotas. Most notably for this paper, Scheduled Tribes and Scheduled Castes are eligible for reservations at government institutions of higher education. Individuals in religions other than Hinduism can also have caste identities. Muslims make up the second largest religious group in India, and have experienced experienced discrimination and persecution. Christians, Sikhs, Jains, and Tribal religions comprise a small minority.

The intersection of caste and SEO is particularly consequential to political debate. Detractors of caste-based quotas in higher education argue that caste's impact on SEO has waned in modern India, and therefore quotas are benefitting well-off ST/SC students to the detriment of poor students from other caste groups. However, testing this claim has been limited due to the general lack of data on student's SEO. Nevertheless, The Constitution (124th Amendment) Bill 2019 allowed for a 10% quota for "economically weaker" students from upper castes at government institutions. A cutoff on family's annual income defines "economically weaker". However, measuring annual income is not straightforward nor are data easy to access. Back of the envelope calculations using IHDS data suggest that 98% of the population qualifies as "economically weaker" according to this cutoff (Deshpande and Ramachandran 2019). While students in this paper's sample were enrolled prior to this new quota policy, these debates highlight the pervasive lack of data on intersecting student identities and the ways this confusion hampers public policy.

5.3.1 Identity Traits in Engineering Student and Age Cohort Population

The AICTE regulatory body provides a few top-level statistics on the engineering student body. I compare these statistics from 2019 which those of the general age cohort in the IHDS in Table 6. It is worth noting that overall, ST/SC students are underrepresented, despite quotas, as are Muslim students who are not explicitly identified in AICTE statistics.

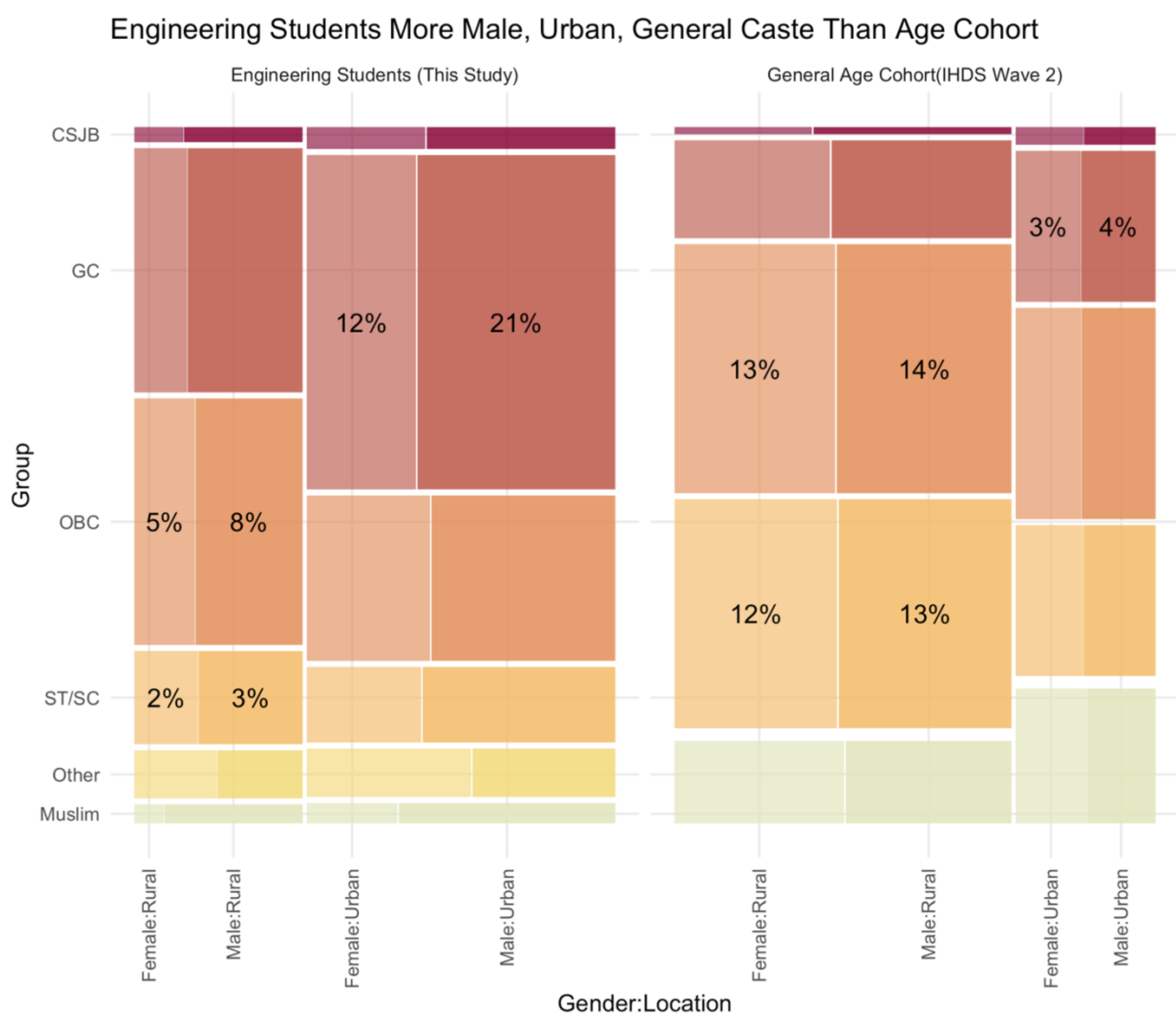
Table 6 AICTE vs. General Age Cohort

Trait		Engineering Admin (AICTE)	Age Cohort (IHDS)
Gender	Female	29%	50%
Religion	"Minority Religion"	7%	20%
	Muslim	Not Provided	15%
Caste	ST/SC	15%	30%
	OBC	35%	36%

Yet these top-level descriptions only examine one trait at a time. In the following Figure 7 mosaic chart, I use this paper's engineering sample to examine the intersection of these three identity variables in comparison to the general age cohort. It's important to note that this sample is not a perfect representation on the engineering student body, yet the comparison provides useful insight on the intersection of these traits. For ease of interpretation, I combine caste and religion into one variable group, following the IHDS's GROUPS variable construction.

The left-hand square represents my sample, while the right is from the general age cohort. First, notice the red, second from the top rectangle on the right-hand side of each population square. This represents General Caste, urban, male students. This group comprises 21% of the engineering students, but only 4% of the age cohort. Interestingly, this sample also finds rural ST/SC/OBC groups underrepresented compared to the population, while their urban caste group peers are represented at about the population proportion. This gap appears particularly severe for rural ST/SC young men and women. Muslim students are consistently underrepresented among the engineering sample. Women are less represented than their male peers in each category, however, women are more like their male peers within caste and religious and location groups than to women in other groups.

Figure 7 Engineering Students vs General Age Cohort



5.3.2 Social Origins and Intersecting Traits

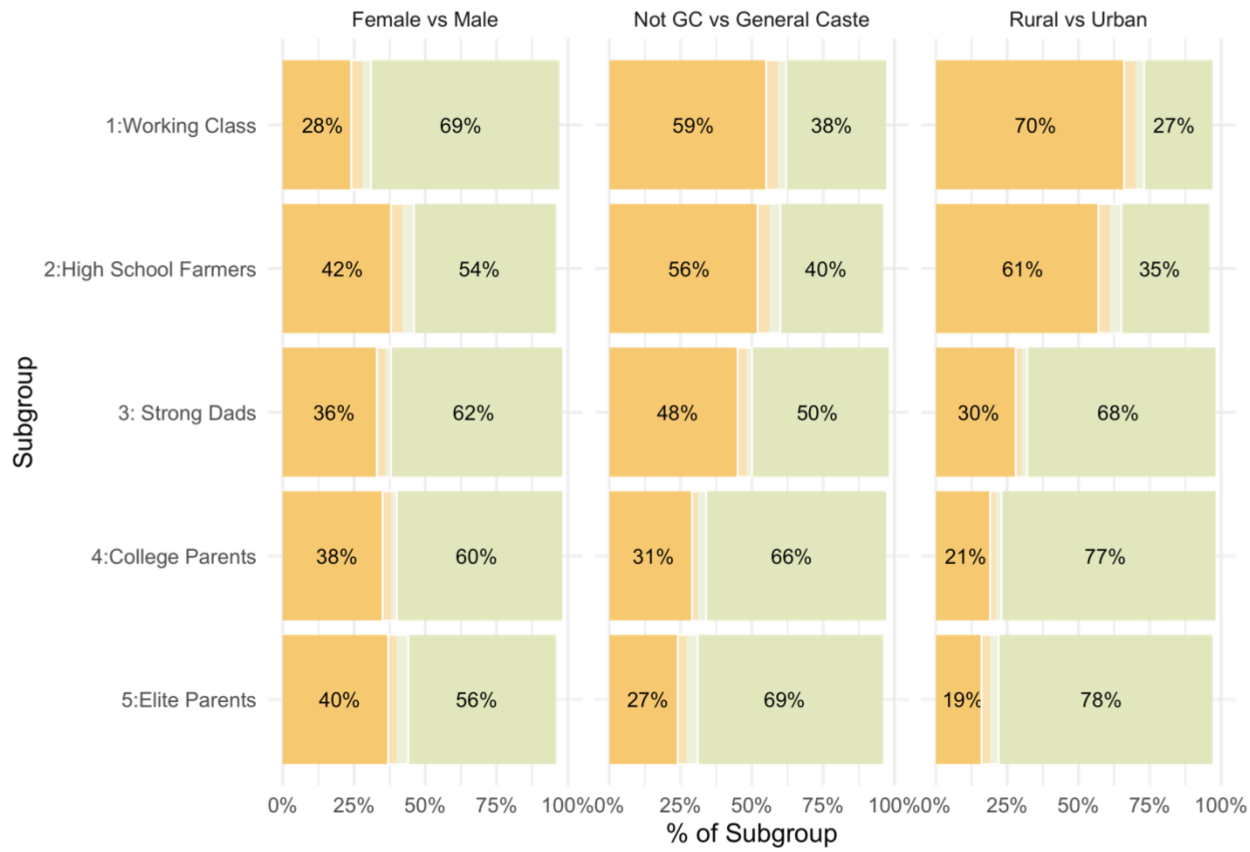
The preceding figure shows that the greater an intersection of disadvantage, the greater the gap between engineering student body and the general age cohort. These intersecting disadvantages persist when considering these traits and SEO in the engineering student body.

I evaluate these competing possibilities using two analytical approaches. One approach estimates conjoint probabilities between SES subgroup membership and each trait. A second examines the extent to which combinations of traits are associated SES subgroup membership using a multinomial logit model with SES subgroup membership as the dependent variable. For ease of interpretation, I make each identity trait into a "dummy" variable, examining proportion female, the proportion "Not General Caste"– that is individuals who are ST, SC, OBC, Muslim or other individuals who do not identify with a caste group, and those living in rural places in their early teens. Both analyses suggest that engineering student intersecting identities more closely align to population disparities. Engineering students from disadvantaged SES social origins are often facing additional challenges from growing up in a remote area, or caste discrimination.

5.3.2.1 Gender, Caste, Religion and Rurality within Socio-Economic Origin Subgroups

For any given student within a given socio-economic subgroup, what is the predicted probability they are a woman? While gender or caste or rurality is not a time-delayed "outcome" of SEO, determining the proportion of women or rural individuals in each subgroup follows a similar analytical logic. I follow Bolck, Croon, and Hagenaars (BCH) approach, updated by more recent work and captured in a SAS procedure provided by the Pennsylvania Methodology Center (Bolck, Croon et al. 2004, Lanza, Tan et al. 2013, Dziak, Bray et al. 2017). This is a "three step" method. The first is estimating the base SEO model, in the prior analytical section. The second step takes the posterior probabilities of belonging to each subgroup for each individual and assigns each individual a subgroup, adjusted with a weight for uncertainty, using the BCH adjustment. The final step estimates Z, the "distal outcome", here, the proportion of individuals in the specified trait group incorporating the step two adjusted weight. I repeat this procedure separately for each identity trait.

Figure X shows the results of these three models, for each subgroup, for engineering students. The darker bar shows the proportion of each socio-economic subgroup who are female, or not general caste, or rural. The small lighter middle bar represents a 95% confidence interval for this estimate, related to the uncertainty around true latent subgroup membership. The percent label, shifted slightly left, is estimated percent in that identity -the center of the confidence interval. The light bar represents the inverse, either proportion male, General Caste, or proportion urban. This analysis shows that in subgroups 1-3 around half of students are not general caste. Conversely subgroups 4 and 5 are 70% General Caste, and about 30% all other groups. The starkest difference between subgroups emerges in comparing rurality. Groups 1:2 and mostly rural, while groups 3-5 are increasingly more urban. Gender differences are more limited. With the exception of 1:Working Class which is only a quarter women, the other four groups are 36-42% women.



In some ways, these patterns are not surprising. Caste and rurality are inextricably linked with socio-economic measures in the general population. They are perhaps a bit surprising given the elite nature of engineering schools. Most lower SEO students navigated their way to engineering college while also confronting generally poorer quality infrastructure in rural areas, and/or caste-based discrimination[4].

5.3.2.2 Socio-Economic Origin Subgroups with Gender, Caste, Religion and Rurality

Alternatively, I consider the socio-economic subgroup distribution within each trait group. I model the three traits as binary covariates in a logistic regression with socio-economic origins as the outcome. Subgroup 5:EP serves as the reference category. This model provides a descriptive summary, not a causal estimate.

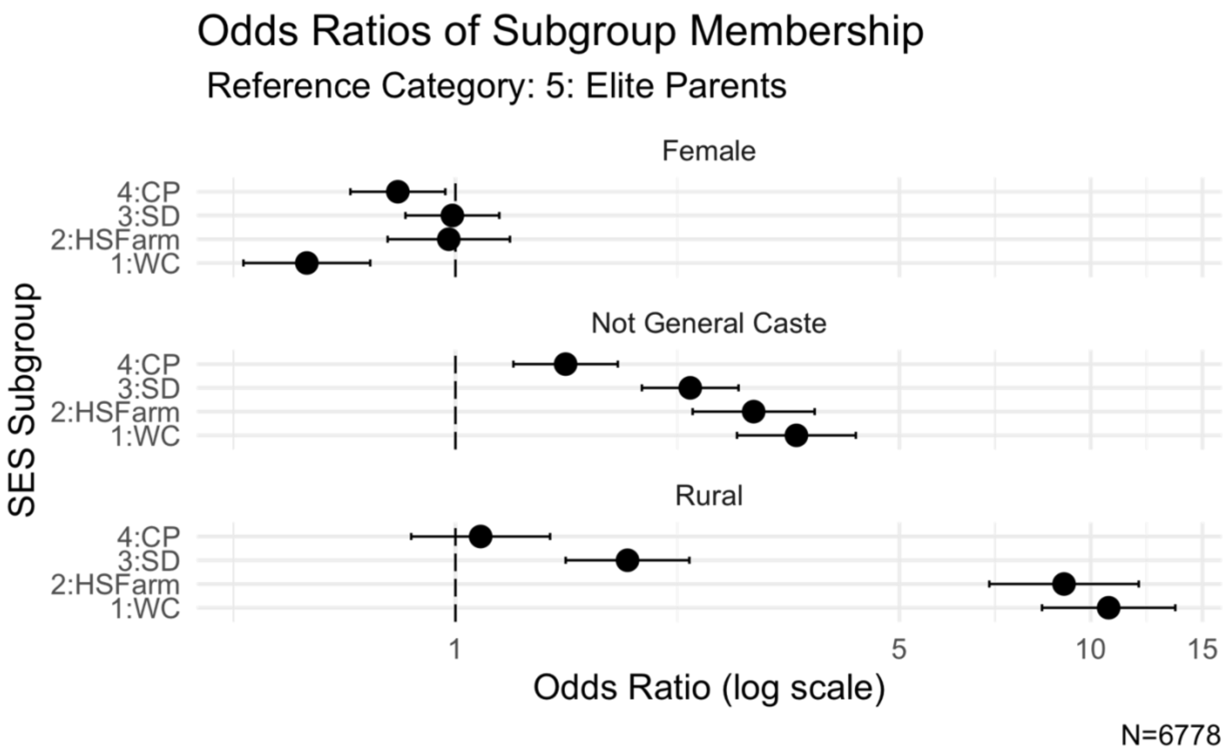
$$log[\pi_{givenclass}/\pi_{5:EP})]=\beta_0+\beta_1 female+\beta_2 rural+\beta_3 notGC$$

The figure below displays the model results in terms of odds ratios. Odds ratios are the odds of being in a given social origin (on the left hand axis) relative to the reference subgroup, 5:EP. Each point is the estimated odds ratio, with the lines representing a 95% confidence interval. An estimate of 1 as the odds ratio suggests "even odds", or no association between that variable and that class relative to the reference. For example, being female rather than male does not change the odds of

[4] Future drafts will disentangle this aggregate category of "not General Caste." Regrettably, the small number of Muslim students in this sample will likely limit inferences for this group.

being in any subgroup relative to 5:EP, with the exception of 1:WC, where being female reduces the odds of being 1:WC. Conversely, being not general caste or rural strongly increases the odds of being in subgroups 1-3 relative to the 5:EP class. A student from a Not General Caste background has 3.4 times higher odds of being in group 1:WC, and a student from a rural background is over 10 times more likely to in 1:WC.

Figure 8



These findings do not support the creamy layer concern. Belonging to any caste or religious group outside the General Caste category is associated with belonging to lower SEO groups. The association for rural students is even more striking, and this minority group is not well documented in administrative data on engineering students. Taken together, both analyses suggest that students from lower SEO backgrounds are not outliers, instead sharing many of the intersecting minoritized identities present in their general age cohort peers.

5.4 Social Origins and Institutional Quality

Indian engineering institutions can range from some of the best in the world to "fake universities,"- schools so fraudulent the AICTE regulatory body must keep an updated list warning off potential students. In between these extremes, many students will find themselves in middling institutions. Within recognized AICTE institutions, student fees, job placement rates, and median starting salaries can still vary significantly (Ravi, Gupta et al. 2019). What is the association between social origins and institutional quality among engineering students? How do these groups intersect with location, gender, caste and religious identity in terms of college quality? I again use the BCH three step approach, estimating the proportion of individuals in an upper-tier institutions within each socio-economic subgroup of students.

Defining an "upper tier" college is a fraught endeavor. An ideal measure of college quality with respect to upward mobility would focus on undergraduate experiences and job placement, and draw from the entire universe of B.E. and B.Tech institutions. However, due in part to rapid expansion in the number and size of institutions, such rankings are in early stages. I use two measures of institutional quality, the first of which I focus on in the body of the paper Average AMCAT score provided by Aspiring Minds for that institution over time assigned a percentile rank. I define "top tier" status as being in the 85th Percentile of AMCAT institutions. This applies to 269 institutions within the universe of 1,800 institutions in the AMCAT database, and seven out of 32 institutions in my sample. An important note, that while 62% of my respondents come from government or government aided institutions, six out of the seven top tier schools in my sample are government schools, meaning 92% of respondents in top tier schools are in government institutions. While perhaps more pronounced in my sample, elite schools have traditionally been government institutions, particularly IITs and NITs.

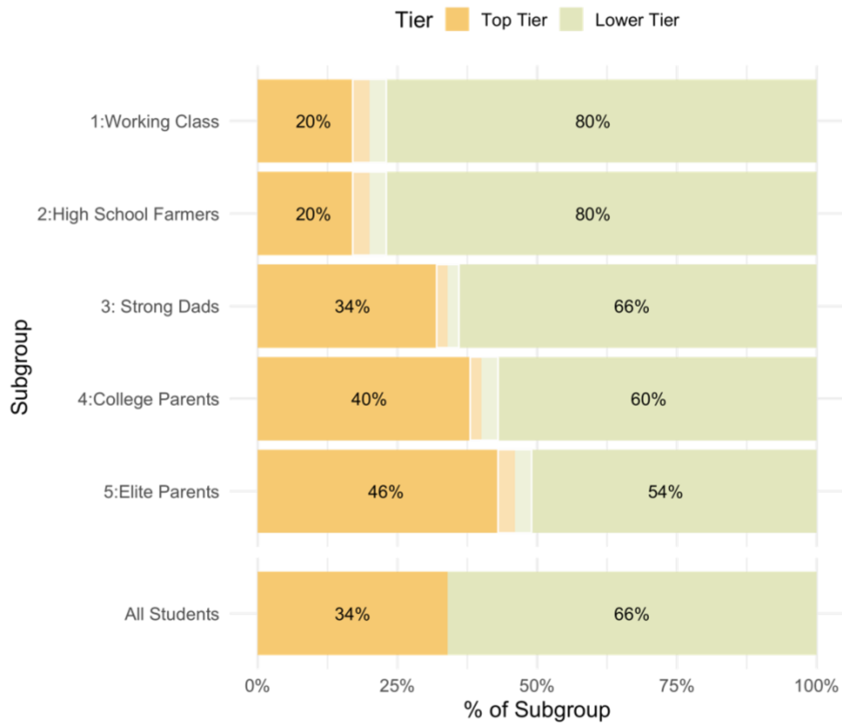
The recent expansion in college seats has overwhelmingly come from new and expanding private institutions. Private schools rely on student fees and attending a private school can be almost twice as costly as a public school. Government aided schools are privately owned, but subsidized by the government. For our purposes, I combine these institutions into a government administered category: these aided institutions resemble pure government schools in their lower student fees and in their need to follow government reservation policies. Private unaided schools instead must rely on higher student fees. Private universities and "deemed to be universities" have met certain standards and have greater autonomy than other institution types

5.4.1 Socio-Economic Origins and Institutional Tier

5.4.1.1 Socio-Economic Origins Alone

Thirty-three percent of my respondents were enrolled in an institution in the top 15% of all AMCAT test taking institutions. I compare each social origin group against this overall average. The predicted probability of being in a top tier institution has a wide range by socio-economic subgroup, even without considering any additional student identities. For students who come from 1:Working Class and 2: High School Farmer backgrounds, the predicted probability of being in a top tier institution is just 20%, meaning the predicted probability of being in lower tier institution is 80%. This is in contrast twith students from the higher SEO groups, who are two times more likely to be in top tier institutions than students in groups 1 and 2. The light color square represents a 95% confidence interval.

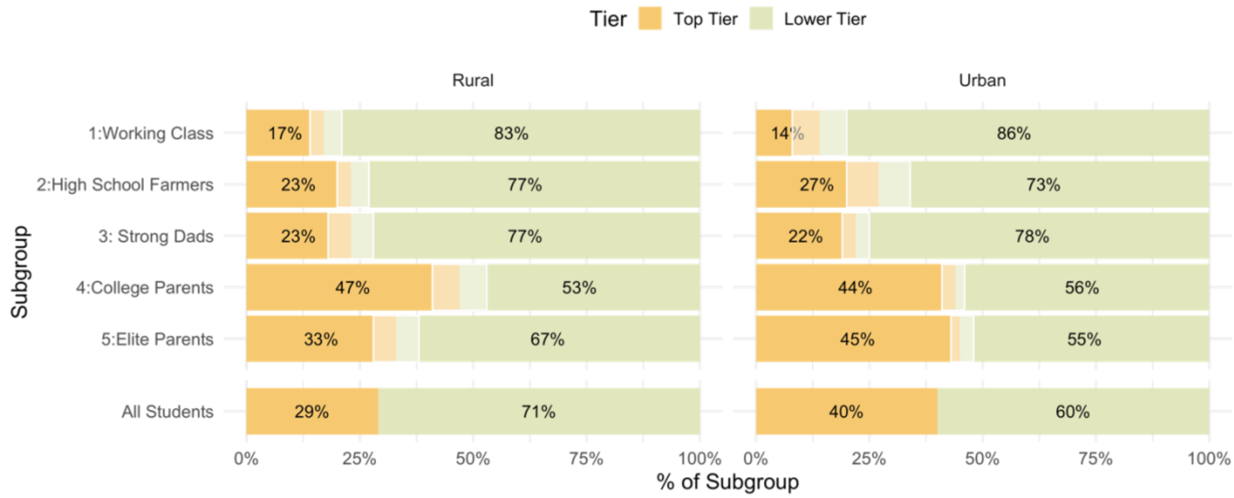
Figure 9 Predicted Proportions Within Tiers



5.4.1.2 Socio-Economic and Rurality of Origins

Without considering SEO, rural students have a lower proportion of students in top tier schools, 29% vs 40%. However, within rural and urban students there is substantial range in predicted probabilities by SEO. Both rural and urban students from lower SEO groups have lower predicted probabilities of being in top tier institutions. The ten percent difference in urban and rural top tier rates may be mostly related to the higher proportion of low SEO students with rural backgrounds. Being urban alone does not appear to offer much higher rates of access to top tier institutions.

Figure 10



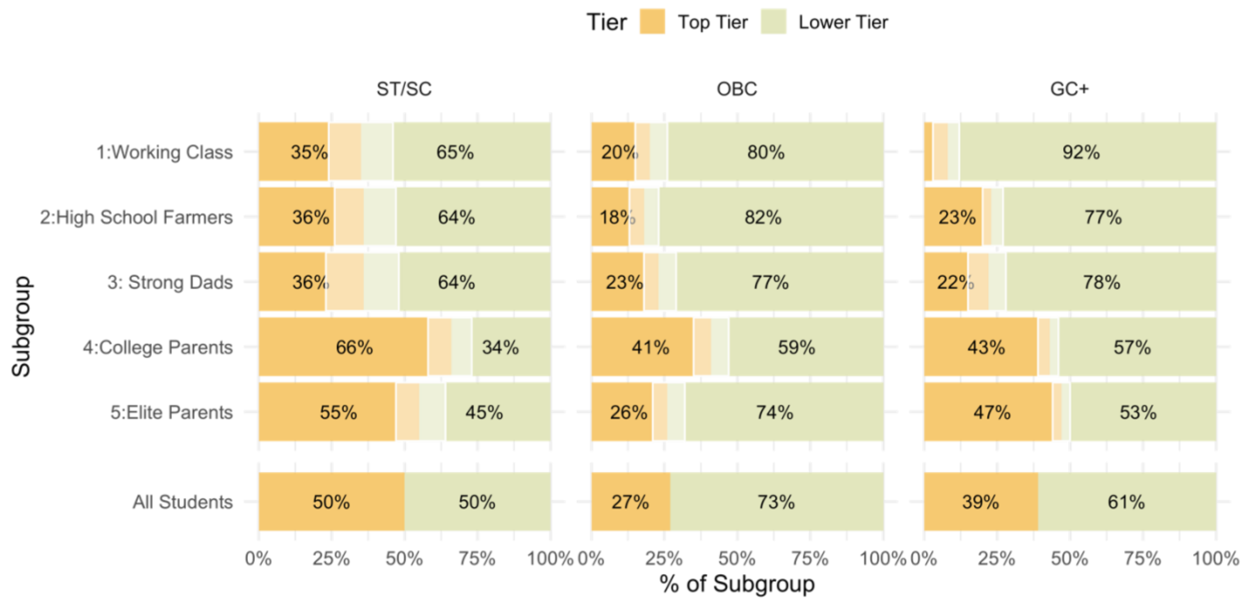
5.4.1.3 Caste

Strikingly, ST/SC students are more likely to be at top tier schools than any other caste group. Even ST/SC students from lower SEO backgrounds find higher representation than their same SEO peers in other caste groups. While few in number, ST/SC students from SEO 4:College Parents, and 5: Elite Parents appear particularly able to navigate to top tier institutions.

OBC students find the lower representation at top tier schools of any caste group, and there is not much difference in the predicted probability of attending top tier schools across all SEO groups. Conversely, there is substantial range within the General Caste student SEO groups. Only 8% of 1: Working Class GC students are in Top Tier schools, in comparison to 20% of GC students in groups 2:HsFarm and 3:SD. Almost half of their more advantaged SEO GC classmates are in top tier institutions.

Muslim students provided too small a sample to reliably estimate differences by SEO group. Given their underrepresentation at every institution type, the hurdle for Muslim students of any socio-economic origin appears to be at enrollment, not tier.

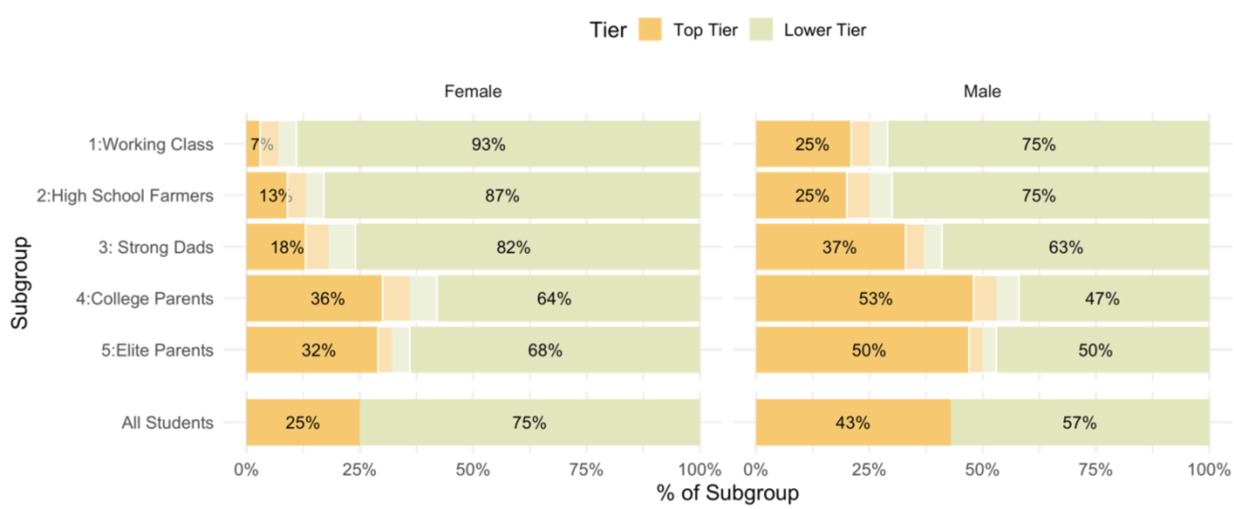
Figure 11



5.4.1.4 Gender

Women students emerge as having some of the lowest predicted probabilities of being in a top tier institution of any group. The disparity in proportion of women who are in top tier institutions is particularly stark for lower SEO group women. While the predicted probabilities follow the same generally lower to higher pattern by social origin subgroup within gender, the predicted probability for women even in the 5: Elite Parent group never exceeds the sample average. This suggests the daughters of mothers who went to college and possibly got a post-secondary degree, and work a salaried job are still no more likely to be in a top tier school than their male counterparts with less educated mothers and more mid-range assets (3:Strong Dads). Similarly, even comparing students from the 1:WC origin, male students are still much more likely to be in a top tier school.

Figure 12



While it's true that women only comprise about 30% of engineering students, and that we found few strong differences between subgroup membership and gender, the above finding is distinct. This finding suggests that the even among women already enrolled, very few make it into the best institutions, even women from the absolute most advantaged families. It's possible that women perhaps are unable to take months or years to study for entrance examinations, yet the average age of women and men in the sample is similar, as is the age of students and top and not top tier schools. Recent work suggests Indian women prioritize safe commutes over institutional quality (Borker 2021).

6 Discussion

This paper provides insight on the emergent patterns of access to an elite destination, engineering undergraduate, during the midst of immense transformation. These findings have implication for both policymakers concerned about equity in education as well as scholars considering social status in times of change.

Through considering multiple dimensions of student's origins, particularly socio-economic origins, this paper adds important nuance to known disparities in access to engineering. While differences in socio-economic access have long been a concern, describing student's complex origins through available data has proven challenging, as their recent debate over the income-based definition of "economically weaker sections" has robustly shown. Using an imperfect but more detailed survey of current engineering students than has previously been available, this paper finds that engineering students tend to be a more privileged population than the general age cohort across five dimensions of socio-economic status. Yet examining the different types of individual's present reveals a substantial portion of students come from backgrounds of disadvantage or mixed advantage.

The interaction between subgroup and other key traits provides further complexity. Rural, low caste, Muslim, and female students are each underrepresented in engineering, separately, and even more severely for individuals with more than one of these identities- e.g rural, ST/SC women. Among the student body, lower SEO groups tend to have higher proportions of rural and lower caste students than higher SEO groups. Being from a rural and/or lower caste background is also strongly associated with belonging to lower SEO groups. On the one hand, this paper finds evidence that the low caste and/or rural students who are successful in entering engineering are not just an otherwise advantaged few. On the other hand, there is clear progress to be made in access for these groups. For example, I find no evidence for the "creamy layer" concern, no large number of socio-economically advantaged ST/SC students. The few ST/SC students present tend to belong to lower SEO groups, and GC students are overrepresented and are associated with higher SEO groups. How these students, and the many lower SEO rural students, marshalled their resources to get to engineering schools is a question for further enquiry.

Slightly different patterns emerge when considering access to the more elite tier. Caste and gender emerge as two important determinants, however the intersection of these traits and SEO is most telling. Focusing on easier to measure disparities in the proportion of women in top versus lower tier institutions misses the even larger gap for lower SEO women. Conversely, ST/SC individuals are more likely to be in top tier institutions, even those in lower SEO positions. This suggests that India's caste-based reservations, which allow lower scoring ST/SC students access to better government schools, works to improve access to better institutions for ST/SC students of all socio-economic backgrounds. While urban students are more likely to be at top tier institutions than rural students, this association appears more related to SEO group, as contingent on SEO group, rural and urban students have similar rates of access.

These differences suggest two distinct processes for getting into engineering institutions generally versus getting into top tier schools. Young people from lower SEO backgrounds are underrepresented in general, and the students who do enroll tend to be at lower tier institutions, particularly for groups with no formal policy supports – specifically lower SEO students outside reservation policies and women. This finding does not necessarily suggest doing away with caste-based reservations to better serve poor students, nor does it necessarily support reservations based on SEO. Targeting to economic need in this context is clearly challenging, as demonstrated by the difficulty in defining "economically weaker sections" by annual income.

Indeed, caste-based reservations only offer limited support, simply a change in how a student's score is considered, at only a portion of engineering institutions. Reservations are not long-term support, they do not provide additional learning opportunities, financial support, nor guidance. The apparent influence of this limited policy in sorting lower SEO ST/SC to top tier schools suggests that despite pervasive inequity in India's education system, there may be other minoritized students at this pathway's later stages, who could benefit from modest supports. Another approach could consider the system of exam preparation, exam scoring, and institutional matching more broadly. The current process, while ostensibly based on raw merit, incurs substantial cost, risk, and distress for students, particularly those with fewer resources and savvy. Aspirants risk failing to match to any institution, or paying large amounts of money for a lower quality education. Limiting costs and low-quality institutions, while expanding access to mid and high tier institutions may reduce some of this risk.

Beyond this specific context, this paper's findings suggest the potential of inductive and multi-dimensional measures of social status measures, particularly in changing places where data coverage is less than ideal. By focusing on types of individuals within the student body, I uncover examples of both extreme upward mobility and persistence, as evidenced by the gaps between 1: Working Class and 5: Elite Parent origins. An individual's intersecting identities can also better inform policy, as it may expose heterogeneity in populations which affects policy impacts, as evidenced broadly by this paper's exploratory description of caste, reservations, institutional quality, and socio-economic origins.

Issues of bias and precision are particularly important when scholars use cross-national comparison to infer which policy environments may support upward mobility, or in decomposing the role of a particular origin variable -e.g. mother's education. However, current data limits progress on these important measurement debates. Even in a future with robust data, social structures will continue to evolve. How often must we update our markers of success, of the power and resources represented by occupations, educational achievements, and other attributes? How can we use the data that are available to make incremental improvements in how we measure concepts in flux? While the field of social stratification and mobility must continue tracking population level dynamics, there remains much to learn in taking a more micro-view of social locations of particular importance- new educational pathways, livelihoods falling by the wayside, or types of young people. Inductive and focused enquiry can help the discipline challenge assumptions, update tools, and better understand the changing world young people must navigate.

Works Cited

Asher, S., P. Novosad and C. Rafkin (2020). Intergenerational Mobility in India: New Methods and Estimates Across Time, Space, and Communities.
Bagde, S., D. Epple and L. Taylor (2016). "Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India." American Economic Review **106**(6): 1495-1521.
Bank, W. (2019). World Development Indicators, World Bank.
Bertrand, M., R. Hanna and S. Mullainathan (2010). "Affirmative action in education: Evidence from engineering college admissions in India." Journal of Public Economics **94**(1): 16-29.
Bloome, D., S. Dyer and X. Zhou (2018). "Educational Inequality, Educational Expansion, and Intergenerational Income Persistence in the United States." American Sociological Review **83**(6): 1215-1253.
Blossfeld, P. N. (2019). "A multidimensional measure of social origin: theoretical perspectives, operationalization and empirical application in the field of educational inequality research." Quality & Quantity **53**(3): 1347-1367.
Bolck, A., M. Croon and J. Hagenaars (2004). "Estimating latent structure models with categorical variables: One-step versus three-step estimators." Political analysis: 3-27.
Bray, B. C., S. T. Lanza and X. Tan (2015). "Eliminating bias in classify-analyze approaches for latent class analysis." Structural equation modeling: a multidisciplinary journal **22**(1): 1-11.
Breen, R. and Jan O. Jonsson (2007). "Explaining Change in Social Fluidity: Educational Equalization and Educational Expansion in Twentieth-Century Sweden." American Journal of Sociology **112**(6): 1775-1810.
Bukodi, E. and J. H. Goldthorpe (2013). "Decomposing 'social origins': The effects of parents' class, status, and education on the educational attainment of their children." European sociological review **29**(5): 1024-1039.
Chabbott, C. and F. O. Ramirez (2000). Development and education. Handbook of the Sociology of Education, Springer**:** 163-187.
Chambers, R. (1995). "Poverty and livelihoods: whose reality counts?" Environment and urbanization **7**(1): 173-204.
Chancel, L. and T. Piketty (2017). "Indian income inequality, 1922-2014: From British Raj to Billionaire Raj?".
Collins, L. M. and S. T. Lanza (2009). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences, John Wiley & Sons.
Dang, H. A. and P. Lanjouw (2018). Inequality trends and dynamics in India. Helsinki, Finland, UNU-WIDER. **2018**.
Desai, S. and V. Kulkarni (2008). "Changing educational inequalities in india in the context of affirmative action." Demography **45**(2): 245-270.
Deshpande, A. and R. Ramachandran (2019). "The 10% Quota: Is Caste Still an Indicator of Backwardness?" Economic and Political Weekly **30**: 27-31.
Dziak, J. J., B. C. Bray and A. T. Wagner (2017). "LCA_Distal_BCH SAS macro users' guide (Version 1.1)."

Emran, M. S., W. Greene and F. Shilpi (2017). "When measure matters: coresidency, truncation bias, and intergenerational mobility in developing countries." Journal of Human Resources: 0216-7737R0211.
Emran, M. S. and F. Shilpi (2015). "Gender, Geography, and Generations: Intergenerational Educational Mobility in Post-Reform India." World Development **72**: 362-380.
Emran, M. S. and F. J. Shilpi (2019). Economic approach to intergenerational mobility: Measures, methods, and challenges in developing countries, WIDER Working Paper.
Frisancho, V. and K. Krishna (2016). "Affirmative action in higher education in India: targeting, catch up, and mismatch." Higher Education **71**(5): 611-649.
Fuller, C. J. and H. Narasimhan (2007). "Information technology professionals and the new-rich middle class in Chennai (Madras)." Modern Asian Studies: 121-150.
Grusky, D. B. (2001). "The Past, Present, and Future of Social Inequality". Social Stratification: Class, Race, & Gender in Sociological Perspective. Boulder, CO, Westview Press: 1-36.
Grusky, D. B. and K. A. Weeden (2008). "Are there social classes? An empirical test of the sociologist's favorite concept." Social class: how does it work: 65-89.
Hannum, E. and C. Buchmann (2005). "Global Educational Expansion and Socio-Economic Development: An Assessment of Findings from the Social Sciences." World Development **33**(3): 333-354.
Hnatkovska, V., A. Lahiri and S. B. Paul (2013). "Breaking the Caste Barrier: Intergenerational Mobility in India." Journal of Human Resources **48**(2): 435-473.
Hout, M. (1984). "Status, Autonomy, and Training in Occupational Mobility." American Journal of Sociology **89**(6): 1379-1409.
Hout, M. (1988). "More Universalism, Less Structural Mobility: The American Occupational Structure in the 1980s." American Journal of Sociology **93**(6): 1358-1400.
Hout, M. (2015). "A summary of what we know about social mobility." The ANNALS of the American Academy of Political and Social Science **657**(1): 27-36.
International Labour Office (2016). Global Wage Report 2016/17: Wage inequality in the workplace, ILO.
Iversen, V., A. Krishna and K. Sen (2017). Rags to Riches? Intergenerational Occupational Mobility in India. Economic and Political Weekly. Mumbai, Athena Information Solutions Pvt. Ltd.
Iversen, V., A. Krishna and K. Sen (2019). "Beyond Poverty Escapes—Social Mobility in Developing Countries: A Review Article." The World Bank Research Observer **34**(2): 239-273.
Jalan, J. and R. Murgai (2008). "Intergenerational mobility in education in India." Indian Statistical Institute, Delhi.
Jonathan, F., J. David, P. L. Jonathan, S. Timothy and T. Jeffrey (2016). "Inequality and Mobility Using Income, Consumption, and Wealth for the Same Individuals." RSF: The Russell Sage Foundation Journal of the Social Sciences **2**(6): 44-58.
Krishna, A. (2010). One illness away : why people become poor and how they escape poverty. Oxford ; New York, Oxford University Press.
Krishna, A. (2013). "Making It in India: Examining Social Mobility in Three Walks of Life." Economic and Political Weekly **48**(49): 38-49.
Krishna, A. and V. Brihmadesam (2006). "What Does It Take to Become a Software Professional?" Economic and Political Weekly **41**(30): 3307-3314.

Lanza, S. T., X. Tan and B. C. Bray (2013). "Latent Class Analysis With Distal Outcomes: A Flexible Model-Based Approach." Structural Equation Modeling: A Multidisciplinary Journal **20**(1): 1-26.
Laursen, B. and E. Hoff (2006). "Person-centered and variable-centered approaches to longitudinal data." Merrill-Palmer Quarterly (1982-): 377-389.
Montenegro, C. E. and H. A. Patrinos (2014). Comparable estimates of returns to schooling around the world, The World Bank.
Narayan, A., R. Van der Weide, A. Cojocaru, C. Lakner, S. Redaelli, D. G. Mahler, R. G. N. Ramasubbaiah and S. Thewissen (2018). Fair Progress? : Economic Mobility Across Generations Around the World. Equity and Development. Washington, DC, World Bank.
NCEUS (2009). Report on Conditions of Work and Promotion of LIvelihoods in the Unorganized Sector. New Delhi, India, National Commission For Enterprises in the Unorganized Sector.
OECD and I. C. f. R. o. I. E. Relations (2018). Agricultural Policies in India.
Ørberg, J. W. (2018). "Uncomfortable encounters between elite and “shadow education” in India—Indian Institutes of Technology and the Joint Entrance Examination coaching industry." Higher Education **76**(1): 129-144.
Oreopoulos, P. and K. G. Salvanes (2011). "Priceless: The Nonpecuniary Benefits of Schooling." Journal of Economic Perspectives **25**(1): 159-184.
Rains, E. and A. Krishna (2020). "Precarious gains: Social mobility and volatility in urban slums." World Development **132**: 105001.
Ravi, S., N. Gupta and P. Nagaraj (2019). Reviving Higher Education in India, Brookings India.
Savage, M., F. Devine, N. Cunningham, M. Taylor, Y. Li, J. Hjellbrekke, B. Le Roux, S. Friedman and A. Miles (2013). "A New Model of Social Class? Findings from the BBC’s Great British Class Survey Experiment." Sociology **47**(2): 219-250.
Shavit, Y. (2007). Stratification in higher education: A comparative study, Stanford University Press.
Shavit, Y. and H. Blossfeld (1993). Persistent Inequality: Changing Educational Attainment In Thirteen Countries, Avalon Publishing.
Torche, F. (2011). "Is a College Degree Still the Great Equalizer? Intergenerational Mobility across Levels of Schooling in the United States." American Journal of Sociology **117**(3): 763-807.
Torche, F. (2014). "Intergenerational Mobility and Inequality: The Latin American Case." Annual Review of Sociology **40**(1): 619-642.
Torche, F. (2015). "Analyses of intergenerational mobility: An interdisciplinary review." The ANNALS of the American Academy of Political and Social Science **657**(1): 37-62.
Upadhya, C. (2016). "Engineering equality? Education and im/mobility in coastal Andhra Pradesh, India." Contemporary South Asia **24**(3): 242-256.
Vaid, D. (2012). "An Empirical Analysis
The Caste-Class Association in India." Asian Survey **52**(2): 395-422.
Vaid, D. (2016). "Patterns of social mobility and the role of education in India." Contemporary South Asia **24**(3): 285-312.
Walters, P. B. (2000). The Limits of Growth. Handbook of the Sociology of Education. M. T. Hallinan. Boston, MA, Springer US**:** 241-261.
Weeden, K. A. and D. B. Grusky (2005). "The case for a new class map." American Journal of Sociology **111**(1): 141-212.

Weeden, K. A. and D. B. Grusky (2012). "The Three Worlds of Inequality." American Journal of Sociology **117**(6): 1723-1785.
WorldBank (1950-2050). Ed Stats, World Bank.

Appendix

(Available in future drafts)