Equity, Efficiency, and Partnership: Examining Secondary Education in Kerala, India

Sameer Nair-Desai

Submitted in Satisfaction of the Requirements for:

BA (Honors) International Relations & Global Business

School of International Relations

University of Southern California - Dornsife College of Letters, Arts, & Sciences

May 2020

Acknowledgements

First and foremost, I would like to thank my thesis advisors - Professors Joshua Aizenman and Jacques Hymans. Both of them provided invaluable advice and feedback and supported me throughout the year no matter the situation. Professors Simone Schaner and Areendam Chanda also provided valuable insight.

I would also like to thank my parents, Supriya Nair and Gaurav Desai, for encouraging me to step out of my comfort zone and attempt something new. Their own lives as researchers has informed my own work, and I am incredibly grateful for their continued love and support. My cousins, Ashwin Nair and Akhil Nair, also provided their own perspective on Indian education, having grown up in the system and departing for the United States as first-generation immigrants.

Three of my peers at USC; Diviya Gupta, Ethan Caspi, and Sahit Menon have also been incredible assets. Throughout the year, they have discussed the literature and theoretical framings with me, sent me articles and data to explore, and assisted with initial data analysis.

The scholars working in Indian education have guided my path and strengthened my knowledge. The research strands of Professors Karthik Muralidharan and Geeta Gandhi Kingdon have been particularly influential.

Lastly, I thank my cohort of fellow honors thesis students, who inspired me to continue working even in trying circumstances, and to push myself to create something beyond my expectations.

Abstract

Debates over the efficiency and equity of Indian schooling are motivated by conflicting theories regarding the purpose and potential of education as either a private good for elite-led path dependence or a public good for ultra-poor mobility. However, the crucial interplay between educational access and learning outcomes has not been fully explored in India. My hypotheses attempt to bridge the gap between the micro-strands of equity and efficiency in an attempt to provide a snapshot of India's progress in ensuring quality education for all. I utilize rich cross-sectional data from the state of Kerala to examine the relationships between disadvantaged access, school-level inputs, and school type with learning outcomes. I find a negative relationship between disadvantaged enrollment and performance, a positive relationship between private education and learning outcomes, and no clear relationship between inputs and learning outcomes. The magnitudes of the estimates motivate further research.

Table of Contents

List of Figures

List of Tables

Chapter 1

Executive Summary

1.1 Motivation

India hosts the world's fastest growing economy and constitutes its largest democracy. Yet the nation ranks amongst the worst in the world in equality. In 2019, Oxfam reported that the top 1% of Indian residents hold over 73% of the nation's wealth ("Income Inequality Gets Worse", 2019). These gaps in equality are increasingly linked to the nation's institutional context, which is portrayed as inclusive yet is often manipulated by a small share of elites. India's cultural heritage and economic power is unique within the global order and deliver a set of problems unlike any other nation's. However, the Union's dilemmas are also consequential for the rest of the world. India hosts one-fifth of the global youth, which translates to over 1.3 billion young people. As the World Economic Forum notes, this presents both a demographic opportunity and a pressing challenge for the Indian economy and also for the larger global workforce (Kedia et al., 2018).

In particular, India's future links to our own as Americans. The relationship between the U.S. and India has been a defining feature of contemporary American life. Since Indian Independence in 1947, the nation has become one of our closest allies. This has resulted in a large transfer of resources between both nations, including technology and human capital. Since 1980, the number of Indian immigrants to the United States has increased by over two million. Indian Americans constitute the wealthiest minority group in the United States, along with Jewish Americans ("A Singular Population: Indian Immigrants in America", 2017). Navigating India's diverse set of problems thus links to our own future as citizens; the peers we work with, the technologies we develop; and the path we take as a nation. In this thesis, I study a microcosm of Indian political and economic reality through a crucial vector for both equality and repression: education.

1.2 Foundational Theories

A set of foundational theories relating to (in)equality guide the literature in education economics. On the one hand, education may be viewed as a platform for minorities and disadvantaged groups to escape poverty and in this way, promote inter-generational mobility. On the other hand, education can be seen as a tool developed and instituted by the elite to perpetuate existing inequalities and ensure that birth endowments dictate life outcomes. Systems of education around the world are diverse mixtures of micro, meso, and macro elements defined by the institutional setting of the time period. The German system promotes the filtration of elite talent from a young age by identifying the most gifted (and often, the most endowed) and placing them in positions to succeed. Meanwhile, the American system prides itself on inclusivity and properly allocating talent; it believes that even the most disadvantaged minorities in America may succeed without birth endowments. Neither system, however, stimulates the *systemic promotion* of minority groups, but rather, identifies specific individuals amongst the disadvantaged to empower.

1.3 Central Hypotheses

In the context of India, which leans towards the German system of filtration, the questions of interest are defined by this structural tension. First, given the recent Union policies promoting universal access, how inclusive is Indian education of disadvantaged minorities? Second, conditional on access, how do they perform relative to their peers? And third, how does performance link to unique external features? Naturally, performance links to an individual's own ability, genetics, and institutional context (at the family, neighborhood, and school level). The literature attempts to control for these (often unobserved) features through randomized experiments and micro-level observational data. I examine two related strands within the literature; which link school inputs (funding, infrastructure, pedagogy) and school type (private schooling as opposed to public schooling) to learning outcomes. Guided by the literature, I form three central null hypotheses of interest:

H_1 = *There is no relationship between* ***access*** *and* ***learning outcomes.***

H_2 = *There is no relationship between* ***inputs*** *and* ***learning outcomes.***

H_3 = *There is no relationship between* ***school type*** *and* ***learning outcomes.***

1.4 Empirical Setting

India is composed of twenty-eight states; each of which are semi-autonomous from the Union. Each state differs in its language, religion, food, and cultural heritage. Due to the heterogeneous nature of the Union, individual studies generally restrict themselves to a detailed study of a single state context. While much of the literature has focused on the worst performers, there exists no detailed study which examines the "best case" of Indian education: the coastal state of Kerala in South India. Kerala has been acknowledged as a leader in the Indian context by most scholars, and overperforms in education relative to other features of state development. If, in even the best case of Indian education, I observe negative relationships between disadvantaged access and learning outcomes; then the results warrant a closer examination of *all* states, rather than simply the worst performers. On the other hand, if Kerala truly outperforms other Indian states, its state policies and model of development can serve as an example for others.

1.5 Data

In order to test my hypotheses of interest, I scrape state standardized test scores for Kerala's Secondary School Leaving Certificate (SSLC) examination. The SSLC, administered at the end of the 10th Standard (Grade), is one of the single most important assessments within a system largely structured around standardized testing. A student's results from the SSLC determines their eligibility and placement for higher studies (in higher secondary school and university). The literature suggests that by the time these tests have been administered, learning gaps within classrooms and between students of various social backgrounds are pronounced. Endowments from a child's neighborhood, family, or peers are far more influential as their educational path progresses. Once a student falls behind the curricula, it is nearly impossible to catch up. Thus, the performance of disadvantaged minorities in these assessments is of particular interest, given their relevance and the presence of learning gaps amongst students. I link these scores to a set of school report cards, which provide detailed information on the enrollments of various minorities (girls; Muslims; Other Backwards Castes; Scheduled Castes; and Scheduled Tribes) within each school. This allows me to examine the relationship between enrollment counts or shares of these social groups and a school's relative test scores.

The report cards also provide information on school level inputs and pedagogy, including the amount of funding received through government grants; the number of computers at a school; whether the school has a playground; and the number of teachers with professional qualifications. Lastly, I am able to examine the school "type." Within India, there are government schools, fully funded by the state and Union; aided schools, which are similar to hybrid charter schools owned by private management but funded by the Union; and un-aided private schools, which are entirely independent private schools. Using these additional covariates, I am able to examine the relationship between a school's input and type with test scores. I analyze 2,702 schools, which represents roughly 57% of the entire universe of registered secondary schools in the state.

1.6 Estimation Strategy

My estimation strategy includes three different regression models, which account for potential heteroskedasticity and multicollinearity within the data. I first run a simple OLS regression with district controls and clustered standard errors (at the school-level). To control for omitted variable bias, I include a range of controls at the district and school level. This regression also includes a series of interaction terms, which help me interpret the interplay of various school features and their relationship to scores (i.e. how does a Private school with a majority Muslim enrollment perform relative to an all-girls Government school?). My second estimation strategy re-frames the simple OLS through a fixed effects regression, to control for a school's location within one of Kerala's unique fourteen districts. Lastly, I run principal component analysis to collapse a larger range of district-level covariates and correct for minor collinearity in the fixed effects model.

1.7 Results

The results suggest a negative relationship between disadvantaged enrollment and performance for minority castes and religions, a positive relationship between private education and learning outcomes, and no clear relationship between inputs and learning outcomes. Access appears to have increased for certain disadvantaged groups, across all school types. There are high enrollments of Other Backwards Caste (OBC) and Muslim students in private schools, suggesting they are more affordable and inclusive than much of the literature suggests. Girls, too, constitute a large share of enrollments across schools, suggesting that unlike elsewhere in India, parents are able and willing to invest in their daughters' education.

On the other hand, across all specifications, schools with higher enrollments of Scheduled Caste (SC) and Scheduled Tribe (ST) students perform worse than those with students from the General Caste. There is no pronounced effect for schools which host large enrollments of OBC students, likely because although this group constitutes a historically disadvantaged minority, they make up a large proportion of Kerala's population (over 50% in many districts). There also appears to be a negative relationship between a school's Muslim enrollment and test scores. However, schools with higher shares of girls tend to perform better on the SSLC, suggesting that gender gaps in education are closing in Kerala. Thus, even in the best case of education, the most disadvantaged groups perform poorly relative to their peers. However, there are clearly heterogeneous effects. It seems that those worst positioned within the Keralan system are ST, SC, and Muslim minorities, which have historically faced the most discrimination across India.

The relationship between inputs and learning outcomes is muddled, reflecting much of the literature on input-based education. While the presence of certain inputs (like the number of computers at a school) seem to relate to higher scores, others (like the amount of state funding received) have no pronounced relationship. However, while the relationship between infrastructural or monetary input investments remains unclear, the results suggest that investments in teachers and pedagogy is likely to improve performance. In particular, the number of teachers who hold professional qualifications at a school holds a strong positive relationship with the school's test scores. The student-classroom ratio also relates to higher scores, which might suggest positive peer effects. Unfortunately, it appears that these relationships are again less pronounced for the disadvantaged. I observe a pronounced negative relationship between a school's test scores and the count of disadvantaged instructors at the school. This poor teacher performance in disadvantaged minorities might be linked to their own lack of training or education, or the regions in which they are teaching (which likely constitute larger shares of students from the same caste or religion). Indeed, schools which consist of a large disadvantaged student base observe positive effects of teachers from the same caste on scores, which suggests positive co-ethnic bias. This suggests the training and recruitment of disadvantaged teachers is particularly important, especially in communities dominated by these groups.

Lastly, the results for school type appear to mirror those from the literature. Un-aided private schools perform better than aided and government schools, despite receiving no state funding. This is likely due to self-selection into these schools, which changes the composition of the student body and thus inflates the average scores due to unobserved individual characteristics. However, aided schools also perform better than government schools, despite receiving less funding. This suggests private management does have some effect on improving scores, even in Kerala, where the state system is renowned.

Although the direction of the coefficients suggest discrimination against the most disadvantaged in Keralan education, the magnitudes do not appear practically significant at scale. A large majority of students sitting the SSLC examination pass and become eligible for higher studies. Based on this result, Kerala seems to be doing a good job in including the disadvantaged and ensuring their success. However, this masks the high dropout rates of disadvantaged groups (particularly SC and ST students) prior to the examination. In most cases, these high dropout rates can be attributed to irreversibly falling behind the curriculum. The simple pass rate alone masks important deviations in the relative performance of those who passed. If these examinations are indeed used as filtration systems to "rank" students relative to one another, then even one tenth of a standard deviation can be the difference between a high school education and admittance to a top university.

1.8 Policy Recommendations and Future Research

Based on the results of this thesis, I propose five programs for Kerala to implement. First, although a brief analysis of the spatial data suggests that the government is allocating funding across districts and regions fairly, more emphasis needs to be placed on areas with large clusters of SC and ST students. Funding should be funneled into these areas to ensure these students do not need to leave their locale to receive quality education. Second, the government should begin to purchase input bundles for schools based on their needs, rather than providing pure cash transfers. In this way, they can focus investment in inputs like computers, which strengthen learning outcomes. The government should also fund teacher recruitment and training programs, and ensure these programs are particularly inclusive of instructors from disadvantaged castes. Third, the government should consider developing a standardized e-learning platform, utilizing technology-aided instruction to deliver low-stakes micro-assessments. These assessments should cater questions to the learning level of the child, based on their performance. This would help schools identify gifted children at a young age, and target those who lag behind the curriculum. Fourth, the government should develop an afterschool tutoring program. Through this initiative, community members (particularly from disadvantaged tribes) could provide additional instruction to students from similar backgrounds who have begun to fall behind the curriculum. Lastly, the government must decide whether to imitate the strategies of private schools to bolster state education (perhaps by developing more exclusive charter schools for the most gifted children) or instead create voucher programs to filter the most talented amongst the poor into private schooling. This paper uses observational data and thus cannot make any causal claims; however, the results should guide further research into the covariates of interest and their links to quality education across India.

Chapter 2

Introduction

My research puzzle relates to a long-standing debate in the field of Indian education, specifically, the lack of consensus on either the relative efficiency or equity of primary and secondary schooling.[1,2] Many scholars have been unable to fully link explanatory variables (i.e. inputs, school choice, and pedagogy) to efficiency (measured through test scores), due to data limitations.[3] They also do not concur on the level of access or equity (often measured as the proportion of disadvantaged students enrolled in a school) afforded to the poor. Even fewer examine the learning outcomes of disadvantaged students *conditional* on their access into systems of education.

Policymakers, too, struggle to integrate both elements of learning. While the Indian government has spent billions of rupees on programs promoting universal access to primary and secondary education and improving infrastructure and funding, serious concerns remain of the quality of government schooling. A number of studies provide evidence that purchasing inputs does little to improve cognitive skills, suggesting the government might be inefficiently allocating their resources to misguided development policies (Hanushek 2003; Das et al. 2011; Borkum et al. 2012; Muralidharan 2013; ASER 2018).

In areas where weakly functioning government schools are prevalent, a new market alternative has emerged (Kremer and Muralidharan 2008). In the last decade, private schools have burgeoned, with lofty promises of stronger learning outcomes and personalized pedagogy (Kingdon 2017). Yet it remains unclear who has access to these fee-based schools, and if they truly deliver better results for their students

[1]For a brief overview of the Indian education system, see Appendix I. A.
[2]For an extensive summary of the recent literature in education economics, see Appendix II.
[3]For a historical derivation of the methodology presented by the literature, see Appendix III. A.

(Härmä 2011; Woodhead et al. 2013). These disagreements reflect a broader theoretical debate on the purpose and promise of education in India. Is education a lens through which the elite perpetuate their own worldview and reinforce their power? Or is it instead a platform for the disadvantaged to achieve socioeconomic mobility?

This thesis measures the efficiency of schools and social castes through standardized test scores, as linked to their equity through enrollment shares (particularly for disadvantaged minorities). The data allows me to investigate how school characteristics link to the overall performance of these minority groups. My hypotheses examine the following linked questions of interest: **(1)** Do schools with higher enrollment shares of disadvantaged students perform better or worse? and **(2)** Do schools with stronger or weaker inputs perform better or worse? These questions relate to the larger central hypothesis of interest, which motivates the research puzzle: is there a pronounced relationship between access or resources (i.e. enrollment shares of the disadvantaged; higher input shares; school type) and performance (test scores)?

I focus my quantitative analyses on Kerala as the one Indian state that typically invites the most praise for its educational achievements and provides strong funding for education. Kerala has placed a historical emphasis on human capital accumulation, and over-performs in education relative to other metrics of state development (ASER 2018; SEQI 2019; Rathore and Upasak 2018).[4]

I exploit extensive data readily available for non-experimental study, scraped from national and state resources.[5] After constructing the theoretical underpinnings of the debate on equity and efficiency in Chapters 3 and 4, the literature review focuses on two subsets of the efficiency vs. equity debate, which inform my hypotheses of interest in Chapter 5, motivate data collection and descriptive analysis in Chapters 6-9, and develop the testable empirical models in Chapter 10.

[4] For an additional justification for the selection of Kerala, see Chapter 6.

[5] Over the course of the last decade, development economics has gained empirical value through a "credibility revolution," driven by the establishment of experimental methods for policy analysis. ("EI Dialogues with Karthik Muralidharan" 2019). Increasingly, a focus on empiricism and causality dominate the field ("EI Dialogues with Karthik Muralidharan" 2019). However, there is value in parametric analysis when RCTs are costly or infeasible (Evers 2017).

Chapter 3

Literature Review

In their book Education and Inequality in India: A Classroom View, Manabi Majumdar and Jos Mooij discuss two overarching theories of the purpose and promise of Indian education. One group of scholars view Indian education as a tool for the elites to retain their power in the status quo, a system of inequality inherited from the British colonial system that continued post-independence (1947) until the present day (Majumdar and Mooij 2012). Another set of scholars reject this deterministic view of education as a tool of elite domination, arguing that education instead serves as a platform of liberation for the disadvantaged. These perspectives derive from theoretical framings of education within Marxism, Institutionalism, Functionalism, Rational Action Theory, and Development Economics.

3.1 Structural Marxist Theories of Education as a Tool for Domination

The seminal work by Samuel Bowles and Herbert Gintis in 1976 captures the Structural Marxist approach to education. Through their Correspondence Principle, the two argue that the notion of education as an institution of liberation is inherently incompatible with the long-term competitive structures of global capitalism. Under capitalism, education instead becomes a tool for the replication of hierarchical social subordination as a well as a mechanism to alleviate and suppress class conflict through hidden socialization (Bowles and Gintis 1976). Bowles and Gintis claim that education and capitalism are mutually reinforcing and adaptive to the needs of one another. Industrialists use school as a normative institution in order to produce a docile and skilled workforce for exploitation, adapting their techniques to the shifting tides of social development (Thompson 2017). Only the children of the elite thrive in this environment, which has been designed for them. Barbara Heyns summarizes their conclusion thus:

> . . . Bowles and Gintis explicitly argue that parental status is largely transmitted to children, that schooling provides scant opportunity for betterment, and that the liberal meritocracy based on talent and ability is rhetorical rather than factual. (Heyns 1978, pg. 1001)

Structural Marxists argue that education serves to enable the survival of capitalism, perpetuating the long-run dominance of the bourgeoise even above their short-term interests. Children are molded to fill the roles of their parents: children of the working class are socialized to continue laboring for the children of the bourgeoise. Under Structural Marxism, education becomes a tool for social domination of the bourgeoise elite.

3.2 Institutional Economics and Political Economy: Theories of Elite-Led Path Dependence

Douglass North and Jeffry Frieden's theories of New Institutional Economics and Modern Political Economy as well as Daron Acemoglu and James Robinson's work on postcolonial institutional development reach similar conclusions, but within the framework of institutional economics and modern political economy (Pastor et al. 2000; Acemoglu and Robinson 2013; Harriss et al. 2013). Although they do not take the radical Marxist approach rooted in class conflict or identify capitalism as the sole cause of elite dominance, these scholars do agree that institutions can serve as a tool for the replication of inequalities. Under North and Frieden's rubrics, the organized economic elites overcome the costs of collective action and leverage their financial resources to accumulate political power. This political power is then translated into policy through the incentives of the elite, fostering path dependence and reinforcing their economic power. A vicious cycle of elite capture thus permeates institutional frameworks and erodes participation and opportunities for intergenerational mobility within education and other institutional frameworks (Pastor et al. 2000; Acemoglu and Robinson 2013; Harriss et al. 2013).

Andreas Hadjar and Christiane Gross extend these theories of path dependence to examine differences between Germany and the United States in the educational institutions of primary and secondary schooling. The US, unlike Germany, openly promotes private alternatives to public schooling and accepts the resulting educational inequalities (Hadjar and Gross 2006). Hadjar and Gross argue that these tendencies link to institutionalized path dependence within these disparate international settings, whereby pro-choice and pro-market normative beliefs in the United States motivate private alternatives and are perpetuated. They use this theory to derive an explanatory model of educational inequality, the "macro-meso-micro-model," which links institutional and structural features of the country (macro), school (meso) and individual (micro) to educational returns. In both models, institutional setting is a crucial element of educational outcomes, even ex-ante of enrollment. If the elite dominate the institutions, they dominate the outcomes as well.

3.3 Functionalist Theories of Education as a Means for Social Maintenance

Émile Durkheim, a prominent French Functionalist of the twentieth century, reaches a different conclusion on the role of education. He argues that education promotes societal values and develops crucial skills in the workforce (Thompson 2018). Durkheim posits that heterogenous systems of education reflect the social realities of different nations and time periods and serve to homogenize the individual within larger societal trends (Barnes 1977). For Durkheim, education creates a communal solidarity for the individual and acts as a crucial tool of socialization. He writes,

> Society can survive only if there exists among its members a sufficient degree of homogeneity: education perpetuates and reinforces this homogeneity by fixing in the child from the

> beginning the essential similarities which collective life demands. (Durkheim 1956, pg. 69)

Thus, education becomes a necessary tool within the capitalist division of labor, creating specialized roles for individuals and instilling a sense of collective purpose above and beyond individual needs. Under Functionalism, education may be a tool for domination, but it represents a tool of social domination over the individual and not necessarily elite domination over the disadvantaged. It is not predetermined that the children of laborers are restricted in their socioeconomic mobility. Rather, education should prepare them for a specialized role which most benefits their social setting. In this sense, education is a necessary precondition for societal survival, even if it might lead to privileging certain social groups over others.

3.4 Rational Action Theory of Education

In their 1997 paper on educational inequality, Richard Breen and John Goldthorpe extend a model of rational action to explain global class divergence in educational outcomes (Breen and Goldthorpe 1997). They seek to derive a theoretical model which explains both traditional cases of educational divergence as well as outliers (like Sweden and the Netherlands) in which disadvantaged groups have made significant gains in both access and performance. Extending Goldthorpe's individual-level framing of rational choice, the authors argue that families and students make decisions linked to rational cost-benefit analysis, whereby independent actors determine whether or not to attain higher levels or standards of education based on their expected payoffs from participation.[1]

Decisions are based on the probabilities of success (passing examinations) versus the immediate benefits of the labor market (finding a job). Importantly, Goldthorpe and Breen do not dismiss class dynamics, noting that costs of remaining in education are inherently lower for elite groups due to higher expected payoffs in the long-term and more resources in the short-term. However, individuals from subordinated groups do

[1] See Goldthorpe 1996.

have the ability to continue pursuing education and may use the payoff of education to re-position themselves in society. Whether or not they make this decision is dependent on the relative costs and benefits of education. Thus, education can be a platform for liberation, should the cost-benefit matrix for the disadvantaged align with the pursuit of higher levels and standards of educational attainment.

3.5 The Misallocation of Talent and "Lost Einsteins"

In development literature, a series of global studies predict individual lifetime income or profession through human capital accumulation, based on parental income (Bell et al. 2016). Others estimate the dominant influence of institutional settings in determining life outcomes (Milanovic 2015). Generally, these papers extend the role of various social and economic "traps," which restrict socioeconomic mobility. The famous conception of poverty traps has motivated experiments aiming to break pervasive cycles of poverty through large asset transfers, job creation, credit access, and a number of other varied government policies (Bandiera et al. 2017). Information traps, where disadvantaged students are precluded from an opportunity because of a lack of knowledge, technical understanding, or self-valuation, also perpetuate social inequalities. Education has been listed as a key solution to both of these social traps, and as a means to promote a skilled and informed citizenry. Yet many economists posit theories of elite capture, whereby the benefits of development either transmitted through external government policies or internal growth (including through education) are appropriated by the elite.

A complementary strand explores theories of elite capture within labor markets. This mirrors a recent subset of development economics concerned with the misallocation of talent (Ashraf et al. 2018). In institutions designed by the elite and for the elite, talent does not dictate success. Instead, connections, endowments, or the ability to "fit into the system" guarantee future earnings. This produces the classic problem of the "Lost Einsteins" – generations of talented individuals dislocated by their institutional contexts and unable to move through society into their optimal positions of welfare maximization. The misallocation of talent constitutes not only a private cost, but also a social cost. By placing those less deserving in positions of power, and excluding those more deserving through restraints, institutional failures produce both inefficient and inequitable social outcomes. As a result of this misallocation problem, many systems of education are unable to develop a skilled workforce amongst disadvantaged and marginalized members of society.

3.6 The Capability Approach: Education as a Platform for Mobility

Amartya Sen advocates for a "right" to education for all human beings, as a springboard for social and economic mobility (Majumdar 2012). In his Capability Approach, Sen centralizes theories of individual freedoms and liberties as mechanisms for development (Sen and Dreze 1995; Saito 2003). Madoka Saito extends this theory to education, focusing on how Sen's influence has affected educational policy. Sen challenges utilitarian economics of "happiness" and the rational choice approach of "income" as measures of individual well-being by focusing instead on satisfying basic needs.[2] He argues that the first step in ensuring quality education is broadly delivering the rollout of education and mandating universal access. His definition of capabilities includes their relevance to personal freedoms as well as their (indirect) roles in influencing social change and economic production (Saito 2003). Education promotes human capabilities, and in so doing reinforces intrinsic and instrumental norms which influence the pursuit of these freedoms. Saito claims that Sen advocates for universal and mandatory education, which to an extent robs individuals (children) of their immediate autonomy, so as to develop in them capacities and capabilities which will further their autonomy later in life (Saito 2003). So important is

[2]The influence of Sen's Capability Approach is not limited to academia. Alongside the United Nations, Sen co-developed the Human Development Index as a core metric of his Capability Approach. Although this measure of development is imperfect, it has influenced significant shifts in policy considerations away from simple income or welfare considerations, towards broader quantitative framings of individual wants and needs.

the institution of education for individual and societal development that Sen regards its potential as a platform for mobility above the immediate wants and needs of the parents or children who might reject education for the immediate (lower) payoff of the labor market.

3.7 Application of Theoretical Framings to the Context of India

India is soon to host the largest and most diverse population in the world, and yet metrics of income inequality place the nation squarely at the bottom of global indices ("India Ranks 147th in Oxfam World Inequality Index" 2018; SEQI 2019; ASER 2018). Education, as a crucial structure of human capital accumulation, may be particularly susceptible to elite capture. Majumdar and Mooij note that standardization within the curricula streams of material and testing ensures that children of the elites might be more able to leverage personal resources to outperform and outpace the poor.[3] Their observation following classroom visits across Andhra Pradesh is telling. They report,

> Schools are often a mirror for the children of the middle class and elite, who see their dominant worldview reflected in the curriculum, while it is a window frame for the students of non-dominant groups who get a peep into the world of the dominant society. Unfortunately, more often than not, primary schools, through their policies and everyday classroom practices, their curriculum and their textbooks, reproduce the background social inequalities within which the school is embedded. (Majumdar and Mooij, 2012, pg. 2)

Thus, a largely English-speaking, private school-educated, high-fee paying meritocracy perpetuates its privileges and structural benefits, including access to jobs, while those outside, restricted to "peeping" at the established hierarchy rather than entering it, are often excluded or left to compete for the lower-paying, less prestigious

[3]For additional information on curricula and methodology in India, see Appendix A.

jobs. Karthik Muralidharan, a leading education economist, notes that "Curriculum has been designed by highly educated elites and reflects a period of time when there was no expectation of universal primary education" (Muralidharan 2013, pg. 29). Muralidharan classifies Indian education as an elite filtration system, whereby students do not develop cognitive skills, but rather are forced into rote memorization of the dense, sometimes outdated, curriculum. In this setting, government standardized testing does not provide meaningful benchmarks for students or parents, nor does it develop cognitive skills in the general population (Muralidharan 2018; "EI Dialogues with Karthik Muralidharan" 2019). Children of the elite are able to perform well on tests due to private tutoring and resources from home. These students are subsequently sorted into professions and vocations based on their test scores and connections, rather than their talent.

A number of other scholars focus solely on how low rates of marginal access and barriers to entry for the ultra-poor and disadvantaged populations foster inequality. They argue that educational outcomes are influenced ex-ante of enrollment, based on a child's socioeconomic background. Indeed, many argue that later inequalities observed in the labor force begin in the earliest stage of an individual's lifecycle, during pre-primary education. Some studies even trace human capital accumulation back to infancy and genetics, arguing that the earliest stages of life are crucial for later academic achievement (Cesarini and Visscher 2017).[4]

Although a substantial amount of evidence suggests Indian education fosters some degree of inequality, it is also evident that learning outcomes have improved over time (Kingdon 2007; ASER 2018; SEQI 2019). This reality is often used to justify government policies. Consider two landmark policies recently passed by the Indian government. Both of these policies, the Right to Education Act (2008) and Sarva Shiksha Abhiyan (translation, "Education for All") directly reference this notion of

[4]The significance of early childcare education cannot be understated but is not the focus of this paper for two related reasons. First, I am interested in how disadvantaged students who have *already* lived within the institutional context of education perform relative to their peers, and second, in order to assess this relative performance, I must exploit some form of standardized assessment unavailable at the pre-primary level.

education as a human right and have pumped billions of rupees into Indian education with the goal of achieving universal primary and secondary education (Woodhead 2013; "Rashtriya Madhyamik Shiksha Abhiyan" 2016). However, many scholars argue that these policies might serve as a smokescreen for the elite within Indian education to maintain their control over the knowledge base (Woodhead et al. 2009; Woodhead et al. 2013; Härmä 2011; Chatterji 2008). It seems unclear from the literature which theories justify the decisions behind education policies and thus properly reflect the realities of the existing education system.

Chapter 4

The Intersections of Equity and Efficiency: Why do we Care?

In testing these respective theories of education, scholars must create metrics for empirical validity. Generally, the literature has responded through constructing indices of "equity" (or access) and "efficiency" (or learning outcomes). Determining whether to start with equity or efficiency might seem like a chicken and egg argument, but the links between these features of schooling cannot be understated. Assumptions about one often lend to assumptions, modelling, and conclusions about the other.

Because of this interchange, it is important to holistically review literature related to both strands of development theory and attempt to integrate both metrics into a single study. Indeed, the circumstances of modern-day Indian education and the theories above merit exploring the intersection between the two. If disadvantaged students gain access to schools which fail to provide them with strong learning outcomes, what's the point? And if a select few make it into elite schools and perform well alongside their wealthy counterparts, what about those who have been left behind? It's unclear whether equity and efficiency have a positive or negative relationship.

This relationship motivates a number of questions for empirical testing. Do higher enrollment shares of disadvantaged groups link to better or worse educational outcomes across different features of institutional quality? How does this relationship vary based on district-level features of income, industry, and literacy? Do school-level inputs make a difference in learning outcomes for the disadvantaged, and do these groups receive more funding to begin with? Based on the theoretical underpinnings of education as a guide, I derive the following null hypothesis:

$H_1 =$ *There is no relationship between* ***access*** *and* ***learning outcomes.***

Through this central hypothesis, I test whether or not the inclusion of disadvantaged students links to better school-level performance. Access relates to equity, measured through the enrollment rates of traditionally disadvantaged students. Learning outcomes relate to efficiency, measured through standardized test scores.

The rest of this thesis proceeds as follows. In Chapter 5, I briefly examine two distinct micro-strands of the literature relating to input-based education and school choice. In Chapter 6, I tie these micro-strands and hypotheses to the literature examining the links between equity and efficiency. In Chapter 7, I construct my relationships of interest and validate Kerala as a setting for my empirical work. In Chapter 8, I review the data and develop my testable empirical models. In Chapter 9, I conduct a preliminary descriptive analysis of the data. In Chapter 10, I review the cross-sectional estimates, and in Chapter 11 I present my conclusions.

Chapter 5

School Features and Performance

Education literature generally defines a set of key relationships which may explain differences in student performance and cognitive development. These are broadly defined within four subcategories of influence: the school, the neighborhood, the family, and the students themselves.[1] Within the context of India, most of these subcategories are unobservable. However, school-level features are readily available for analysis. This influence of the school can be further divided into inputs, pedagogy (which include teaching quality, curricula, and the environment of the classroom), school type (whether a student attends a private or public school), and the peer effects of fellow students. Increasingly, literature which focuses on learning outcomes has shifted to examining the specific links between funding, inputs, and "school quality" (or pedagogy) with cognitive skills development (measured through test scores). My empirical model is developed from these sub-relationships of interest.

5.1 Input-Based Education

A number of studies have found that an increase in funding does not result in better performance, due to household substitutions away from educational spending (Das et al. 2011). Yet other studies demonstrate that depending on where the funding is placed, learning outcomes might improve. For example, funding certain types of input purchases, like libraries, computers, or meal programs might not be effective at improving cognitive skills development (Borkum et al. 2012). However, other inputs, like science laboratories or internet connection could improve cognitive skills by providing a technological foundation for instruction (Banerjee et al. 2005). Using

[1] This thesis focuses on the influence of school-level characteristics but integrates some features of the other three sub-categories where data is available.

funding to improve teacher quality or accountability may also improve performance outcomes, although a consensus has not been formed in the literature (Mbiti et al. 2019).[2] Thus, input-based education provides mixed results. It is unclear whether stronger inputs truly fuel stronger outcomes, and if so, which inputs best complement student learning.

The debate over input-based education is particularly salient considering the massive government spending campaigns on public education, and India's focus on improving input-based metrics of educational quality ("Private vs. Public Government Schools" 2019). Because of the Indian government's large financial commitment to educational improvement, groups are divided on how and where the new money should be spent. Government policy has traditionally aligned itself with an input-focused strategy of educational development, but many academics contest this approach as an inefficient use of resources which does not improve student learning. It is important to examine this micro-strand not only due to its policy implications for student learning, but also for considerations of equity. If evidence suggests that funding stronger inputs has improved performance, it behooves policymakers and academics to ensure access to these inputs across disadvantaged groups.

Changes in student performance are largely conditional on household responses to additional funding, the utility of the input to the student, and the proper employment of the inputs by schools and teachers. In 2003 Eric Hanushek, a seminal author in education economics, published a paper titled "The Failure of Input-Based Schooling Policies" which examined the inefficiencies of global policies focusing on school inputs rather than incentives (Hanushek 2003).[3] Using historical evidence from the United States (1969-1999), Hanushek illustrated a drastic decrease in pupil-teacher ratios and increase in spending.

Hanushek argued that the reduction in class sizes and additional expenditure were

[2] Other studies by Muralidharan provide some validation in the Indian context, suggesting that government teachers generally perform poorly, are frequently absent, and do not teach to the level of their students unless incentivized to do so monetarily or through the threat of punishment (see, for example, Muralidharan and Sundararaman 2013).

[3] See Appendix III. A. for Hanushek's contribution to the empirical methodology.

not supported with stronger test scores. In fact, test results under the National Assessment of Educational Progress steadily declined in most subject areas over the same period. While data limitations prevented causal interpretation in international cross-comparative analyses, Hanushek also reviewed the available international test data and reached similar conclusions. After presenting historical evidence, Hanushek ran empirical regression analysis comparing international test scores to spending. Despite differing institutional frameworks, there were almost no significant relationships between funding and performance across his nations of interest. However, when examining developing nations, Hanushek found preliminary evidence that resources may improve outcomes, and attributes this to their lower initial level of development and need for baseline funding. Hanushek's paper could not establish causality, but motivated research across international contexts in microeconomics.

Other scholars have begun to examine if certain inputs might improve outcomes or if the mechanism by which the inputs are bundled makes a substantive difference. Isaac Mbiti and his co-authors examine an experiment run across 350 schools in ten Tanzanian districts, to examine complementarities between funding and incentives. The authors theorize that the observed insignificant relationships between inputs and learning outcomes might not link to the inputs themselves, but rather to low levels of ability or effort by the teachers utilizing the inputs (Mbiti et al. 2019). At the same time, the relationship between effort and inputs might be reversely correlated. For example, perhaps teachers exert low effort because they face poor inputs and are unable to adequately teach given their difficult environments.

The experiment worked with a local nonprofit to randomly allocate schools into four different groups: one which received a funding grant, one which received teacher incentives[4], one which received both, and a control. Students were then given a "low-stakes" test and a "high-stakes" test to determine the change in learning outcomes. The authors find that in schools with grants, household expenditures fell, and student learning outcomes remained unchanged. This aligns with the prior re-

[4]Based on prior work by Muralidharan and Sundararam in their 2009 experiment.

search from Jishnu Das and his co-authors (Das et. al 2011). In the second case, student scores improve insignificantly in the low-stake tests but substantially in the high-stakes test. However, in the third case, where schools received both grants and teacher incentive programs, students scored substantially higher in both tests. The authors find further evidence of complementarities between the two programs on longer timescales, after re-examining student learning outcomes two years later. Thus, the authors conclude that complementarities might exist between bundles of inputs (i.e. funding for infrastructure and training or incentives for teachers).

Based on the academic literature, we might expect the general relationship between inputs and learning outcomes to mirror the null (no significant correlation). Specific inputs or bundles of inputs, however, might reveal a positive relationship. But despite this mixed evidence, Indian policymakers seem to have reached their own conclusions. Recent legislation has focused primarily on funding improvements in infrastructure, rather than improvements in learning outcomes.

According to Brookings, the Indian government has allocated nearly $14 billion towards education for the 2019-2020 fiscal year alone (Gustafsson-Wright and Boggild-Jones 2019). The majority of this money will be spent on financing for school development, including the provision of toilets and libraries, the subsidization of training programs for government teachers, and free meals for students. The government has implemented policies promoting positive discrimination in favor of excluded castes, but it remains unclear how successful they have been. If disadvantaged students are enrolled in schools with weaker inputs and this links to poorer learning outcomes, they suffer from exclusion in both dimensions. Examining the schools in which underrepresented castes are enrolled thus merits not only considerations of performance, but equity. These questions motivate one of my primary relationships of interest, which I examine within the context of disadvantaged access.

H_2 = *There is no relationship between* ***inputs*** *and* ***learning outcomes.***

5.2 School Choice

Another interesting trend in Indian education has been the establishment of a growing number of unaided private schools, as opposed to government schools and aided private schools (Härmä 2011). Unlike aided and government schools, unaided private schools are managed and operated by private individuals as fee-charging institutions.[5] A number of recent studies suggest that unaided private schools increasingly account for larger proportions of student enrollment in both urban and rural settings (Chatterji 2008; Muralidharan 2013; Majumdar 2012; French and Kingdon 2010; Kingdon 2017). Geeta Gandhi Kingdon claims that nearly 96% of the increase in urban primary enrollment from 1993-2002 was in private schooling, while the Hindustan Times reports that between 2010-11 and 2015-16 enrollment in government schools fell by 13 million but grew in private schools by 17.5 million (Saha 2017).[6]

Because a large number of private schools are unrecognized, and additionally because government offices are incentivized to inflate enrollment in government schools, these statistics are actually substantial underestimates of true growth in the private sector (Kingdon 2017). So, arguments about access clearly have to be recalibrated to match current realities in Indian society. Scholars are interested in whether this new "low-fee" private sector can deliver better learning outcomes for the poor or are instead structures to facilitate "elite flight" away from overcrowded public education. The theory of elite flight suggests that parents themselves care about this interplay between access and performance. Rich families might not directly state their fears of

[5] Note that for all intents and purposes, aided schools are the same as government schools. Henceforth, "unaided" schools will be referred to as "private," and "government" and "aided" schools as "public." A number of studies make this same grouping, even referencing aided schools as "government aided."

[6] It is important to note that despite this growth trend, the majority of children (65%) continue to learn in government schools, and state funding is funneled into public education (Saha 2017). Selected statistics on the growth in private schooling across Indian states are provided in Appendix A.

an "equitable" (accessible) system linking to lower "efficiency" (school test scores), but elite flight would reveal their preferences.

If the rich and powerful are fleeing public education because of the recent inclusion of the disadvantaged, clearly, they believe that higher enrollment rates of disadvantaged students link to poorer outcomes for their own rich children. If this "elite flight" links to negative spillovers, which results in lower efficiency for the poor who remain "trapped" in public education, then even the poor amongst India would be concerned with how access can link to learning outcomes within the current system. On the other hand, those who posit that Indian education can be a springboard for social and economic mobility argue that poor students should have access to education (hence the recent policies of the Indian government), but should also be supported in order to succeed (thus the various policy recommendations of scholars to supplant schooling with remedial education and reform curricula away from the lens of the elite).[7]

The Right to Education Act and Sarva Shiksha Abhiyan promised citizens free and universal primary education for all, presenting education as a "right" for the poorest of India. The Rashtriya Madhyamik Shiksha Abhiyan extended this right to secondary schools across the nation ("Rashtriya Madhyamik Shiksha Abhiyan" 2016). But while the Right to Education Act and its associated programs are certainly milestones in India's education policy, the literature suggests that they have not fully delivered on their promises of universal access and quality.

Joanna Härmä reports on students in low-fee private schools in rural Uttar Pradesh, using evidence from a household survey data of 250 families across 13 villages in the J.P. Nagar district. Data was collected on occupation, income, assets, and family information. Most families in the villages earn their income from subsistence agri-

[7]While I do not present the detailed findings of historical and policy-specific work external to my hypotheses, an immersive overview of the literature can be found in Muralidharan's book chapter on the priorities for Indian education policy, Sonia Bhalotra and Bernarda Zamora's working paper on primary education, K. Biswal's overview of secondary education, and Kingdon's discussion paper on private schooling (Muralidharan 2013; Bhalotra and Zamora 2008; Biswal 2011; Kingdon 2017).

culture. Härmä splits her analysis into three related questions. First, which type of school do parents prefer? Next, what type of school do most students end up attending? And finally, does access to school type link to income? She finds that almost all parents prefer private schools to public schools, but despite this preference, most students are enrolled in public government schools.

Härmä uses logit regression modelling, constructing an asset index based on familial poverty to derive the explained variable of school choice. She finds that the larger a family, the less likely that a child attends a private school. Christians and Muslims as well as so-called Scheduled Caste and Scheduled Tribe students are also less likely to attend private school. The lowest two quantiles by income have very low rates of access to private schools in Härmä's setting. She concludes that rural private schools are not pro-poor, but rather increasingly reflect the "elite flight" of the higher socioeconomic castes away from the government sector (Härmä 2011). However, another group of scholars posit that low-fee private schools offer the poor a path for liberation.

Kingdon's 2017 discussion paper examines low-fee private schools using data from the Unified District System for Education (U-DISE), National Sample Surveys (NSS), and Annual Survey on Education Report (ASER), to compare private schools' size, growth, salaries, per-pupil costs, pupil achievement levels, and cost-effectiveness as compared to public schools. Kingdon uses this set of survey data and her own calculations to present two crucial findings. First, the better the government schools in the region, the less the demand for alternative private modes of education. Next, the fees of the few private schools that survive in these areas are high (due to limited supply). However, in states with low quality public schooling, the demand for private schools is strong, more of them are established to meet this demand, and fees are lower because of competition.

Kingdon concludes that the literature has underestimated the level of access which ultra-poor households have to low-fee private schools. Regardless of the quality of public schooling and associated size of the private sector, this data suggests that a good proportion of private schooling may actually be affordable to the ultra-poor. Kingdon notes, "This evidence discredits the oft-repeated belief that much of private schooling in India is elite and exclusive" (Kingdon 2017, pg. 23).

Disadvantaged access to private schooling is particularly salient because much of the prior evidence on the relative efficiency of private and public schools suggests that private schools deliver better learning outcomes than public schools.[8] Yet despite a wealth of literature documenting a private school advantage, the most recent work in education economics has revealed a more measured private school effect, with a few notable studies claiming no difference in achievement by school type, after correcting for selection bias.[9] A recent local study by Muralidharan and Sundararaman has come closest to establishing a causal link between school type and achievement by correcting for selection bias almost entirely (Muralidharan and Sundararaman 2013). Although the research focused on relative efficiency by school choice, the experiment also increased access for disadvantaged students. This enabled the authors to examine the performance of the schools into which the disadvantaged gained access (although they do not provide performance metrics by caste).

The RCT features a voucher program in Andhra Pradesh, utilizing a two-stage lottery system in order to examine both student-level and market-level outcomes. A key element in the design is the randomization of not just the distribution of vouchers, but also the placement of students into school systems. This allowed the authors to further examine spillover peer effects on (1) public school students who did not win the voucher and remain in public schools, along with (2) private school students who experienced the influx of lottery winners. In other words, they could compare two different treatments and controls: private school students who experienced the inflow of disadvantaged students to those who did not; and public-school students who experienced the outflow of their peers to those who did not. They found no substantial differences in test scores over time by school type, and no significant peer

[8]See Appendix B for a detailed overview of the literature on school choice.

[9]Selection bias refers to the high likelihood that elite students might pre-sort into private schools and be better prepared to succeed due to external household or individual characteristics, thus endogenously (non-randomly) boosting school scores.

effects. Both of these directly contradict the theories of efficiency and equity posited by earlier literature. The first finding suggests that there might not be a private school advantage at all. The second finding notes that the negative spillovers on the ultra-poor trapped in public education, proposed by the "elite flight" theory, might be unfounded (Muralidharan and Sundararaman 2013).

Thus, much of the literature is either inconclusive, fails to control for selection bias, is limited by small sample sizes, does not account for the realities of Indian standardized testing, or does not examine performance by caste, religion, or gender. Questions of access and equity naturally link to those of efficiency, and this does not seem to be fully highlighted in the debate. Even if the ultra-poor might have better access to public education, if, as Majumdar believes, the system of education caters mainly to the elite, those coming from disadvantaged backgrounds may not fully develop their cognitive skills.[10] On the other hand, if private schools produce better learning outcomes but do not include the ultra-poor, they do not produce the universal access Indian policy attempts to prioritize. Even if private schools are "better" than public schools, what's the point if only the elites benefit? Parth J. Shah of India's Centre for Civil Society pointedly summarizes the significance of the debate around school choice...

> If we had convincing evidence that [the] private sector delivers better quality of education at a lower cost, then how much weight and effort do you want to put on improving the state education system? Since we have finite intellectual and advocacy resources, the question I struggle with in our School Choice Campaign is how much do we focus on improving the state system, versus the [current] efforts to liberalize and support the private provision of education? (Muralidharan 2013, pg. 23).

[10]See Muralidharan's conversation with Education Initiatives for more on this.

Shah's statements on school choice link directly to the dilemma of the research puzzle. However, private and public schools are not substitutes, but rather complements. While parents and students certainly must decide which institution to attend based on their resources and perceptions of quality, policymakers must consider both subsectors for legislative reform. The debate should not revolve around private vs. public, but rather consider private and public, navigating integration between both school types. In other words, improving the access and quality of one school type should not replace efforts to supplement the other.

While this thesis attempts to provide an answer to Shah's questions of relative efficiency, its policy recommendations based on these findings do not isolate solely the metric of performance.[11] Rather, the paper examines the relative efficiency of private and public schools to determine just how wide the gaps between provision and quality are across school types which may be more or less accessible to disadvantaged groups. From the literature on school choice, I derive another relationship of interest.

H_3 = *There is no relationship between* ***school type*** *and* ***learning outcomes.***

[11] Note also that Shah's argument rests on the limited resources of the state, and yet many studies on input-based education seem to suggest that India is misplacing the focus of their massive funding campaigns.

Chapter 6

Integrating Equity and Efficiency: Lessons from the Past

Although many studies focus on a particular micro-strand of education policy and either isolate their findings or only examine the interplay of equity and efficiency tangentially, some scholars do attempt to fully integrate both indices. In the third chapter of his book, The Progress of Education in India, Vani Kant Borooah uses rich household-level data from the India Human Development Survey (IHDS) to investigate which features might explain educational gaps amongst social groups. His study uses administered test data from the IHDS nationwide surveys.[1]

Borooah argues that scholars should create "equity sensitive" scores to account for the distributions of achievement within different social groups (Borooah 2017). Borooah identifies six social groups of interest: Scheduled Tribes (ST); Scheduled Castes (SC); Non-Muslim Other Backwards Castes (NMOBC); Muslim Other Backwards Castes (MOBC); Muslim Upper Classes (MUC); and Non-Muslim Upper Classes (NMUC). He further stratifies outcomes based on gender, school type, and household expenditures.

Borooah's results suggest that NMUC students score highest, while ST students score lowest. Predictably, wealthier students outperform poorer students. Private school children also significantly outperform their public-school counterparts. The gaps

[1] India Human Development Survey (IHDS), is a collaborative project between the University of Maryland, the National Council for Applied Economic Research, and the University of Michigan's Inter-University Consortium for Political Science Research. The IHDS is one of the largest panel surveys in the world, consisting of two extensive rounds of data collection from 2004-2005 and 2011-2012. The surveys cover over 40,000 households across India and measure a broad range of socioeconomic features.

between NMUC (elite children) and disadvantaged students remained pronounced across both rounds of the survey. Borooah explains this difference in performance through differences in access, noting that NMUC children were much more likely to attend high-performing private schools. Expenditures on schooling were also highly correlated to expenditures on private tutoring, and additional household resources for elite children.[2]

This gap in educational access links to performance and vice-versa, but also might explain persistent institutional inequalities. As an example, private schools' primary medium of instruction is English, a language highly valued in the labor force. The elite students of the NMUC class were able to afford this instruction, and thus learned the language (as a result, these students had much higher rates of literacy). Due to household resources and "better" schooling, the elite students performed well in their tests. If cognitive skills development links to later job outcomes and thus guarantee higher incomes for the elite, this cycle becomes self-perpetuating.

To correct for intergroup differences in achievement and further validate his findings, Borooah constructs equity-sensitive scores. For each social group, Borooah reduces the mean score by the degree of inequality (distributional differences) of students' test scores within the group. These revised scores were constructed as arithmetic means, geometric means, and harmonic means. Not only were the test scores of disadvantaged groups lower than the elites, but they were also more unevenly distributed. This was particularly the case in writing scores, a skill integral to labor force participation.

In his regression analysis, Borooah runs an ordered logit regression to control for individual and household features between castes. Using this predictive model, Borooah finds NMUC groups are significantly more likely to score higher marks (Borooah 2017). These initial results control for income and external features, and thus look at "identical" students whose only difference is caste membership. Borooah's study sets the stage for new research in "equity-sensitive" education. However, Borooah

[2]Features which are treated as unobservable in this study.

provides a larger picture of Indian education, a "broad stroke." He simultaneously restricts his analysis to survey data which, though rich, is limited in its observable features and may be subject to respondent bias.

Borooah also does not examine outcomes conditional on specific sub-categories of interest. For example, how do Scheduled Caste children with access to a library in their school perform relative to those without? What about the overall performance of the school itself, as opposed to a restrictive analysis of caste outcomes? Lastly, Borooah restricts his study to children ages 8-11, but evidence suggests that as children grow older, learning gaps increase ("EI Dialogues with Karthik Muralidharan" 2019). This thesis will attempt to answer these important supplemental questions. In the following chapter, I crystallize these theoretical and empirical debates in the literature to validate the derivations of my central hypotheses and expand the scope of the research integrating equity and efficiency.

Chapter 7

Validating the Empirical Setting and Reviewing the Hypotheses of Interest

7.1 Validating the Empirical Setting of Kerala

Before proceeding with an exploration of the data and introduction of the empirical model, it is necessary to validate the empirical setting of Kerala. Today, Kerala is widely regarded as a leader in educational achievement and learning outcomes (Asadullah 2012; ASER 2018; SEQI 2019). It is evident that Kerala outperforms in education relative to its income levels (Rathore and Upasak 2018; ASER 2018; SEQI 2019). Kingdon notes that in addition to these strong learning outcomes, inequality in access to secondary education is amongst the lowest in Kerala (Kingdon 2007).

The fact that Kerala has historically performed well relative to India makes it a unique case, and one worth examining in further detail. My primary justification for selecting Kerala is due to the state's strong outcomes in equity and efficiency. If in even the best case of Indian education, there are observable differences in access and learning outcomes across districts and castes, the rejection of my null hypotheses gains added value.

If Kerala has been successful at integrating the disadvantaged, a review of local policies which expand inclusion and performance may serve to inform other states on a future policy direction. A paper by Udayan Rathore and Upasak Das explains this current success through historical institutional reforms. The authors argue that precolonial historical endowments of kingdoms focused on initial education and social policies which were inclusive and equitable, while state activism during the 1950s post-independence bolstered funding for education (Rathore and Upasak 2018). Because Kerala historically pursued egalitarian reforms in land distribution and small market development, first through feudalism and later socialism in the 20th century, the state enjoys greater levels of equality to the present.[1]

Recent trends in Kerala also align well with the context of the puzzle. Between 2005-2006 alone, private school attendance increased by more than 10 percentage points. At the same time, Kerala has dedicated significant resources specifically to improving the quality of government schools and for the first time in 25 years, has seen an extended increase in public school enrollments ("Kerala's Public Schools See Rise in Enrolment for Third Year" 2019).[2] ASER's 2018 study reported that 58% of Keralan children enroll in government or aided secondary and higher secondary schools, compared to 34% in private unaided schools (ASER 2018). This provides strong variation in school type which is well reflected in the data.

Kerala also has a useful demographic makeup for study. There are fourteen unique districts in Kerala: Kasaragod, Kannur, Wayanad, Kozhikode, and Malappuram (North Kerala); Palakkad, Thrissur, Ernakulam, and Idukki (Central Kerala); Thiruvananthapuram, Kollam, Alappuzha, Pathanamthitta, and Kottayam (South Kerala). Analysis can also scale up to the regional level (North, South, and Central), or extend down to the sub-district level (for a total of forty sub-districts).[3]

Around 9% of the population are Scheduled Caste, 1.5% are Scheduled Tribe, and 26.5% are Muslim. The proportion of Scheduled Caste and Scheduled Tribe individuals are lower than the Indian averages, but the literacy rates of these groups are substantially higher. The government has committed roughly $80 million to education for Scheduled Caste and Scheduled Tribe children in this year's budget and are committed to improving performance of marginalized groups ("Kerala Budget Analysis 2018-2019" 2019).

[1] For a theoretical framing of factor endowments and institutional equity, see Pastor, Manuel et al. 2000.

[2] The government spends a significant amount of state resources on funding education and improving school inputs (16% of the 2018/19 budget).

[3] It is often infeasible, however, to gather sub-district level socioeconomic controls.

Figure 7.1: Scheduled Caste Population by District (Proportion)

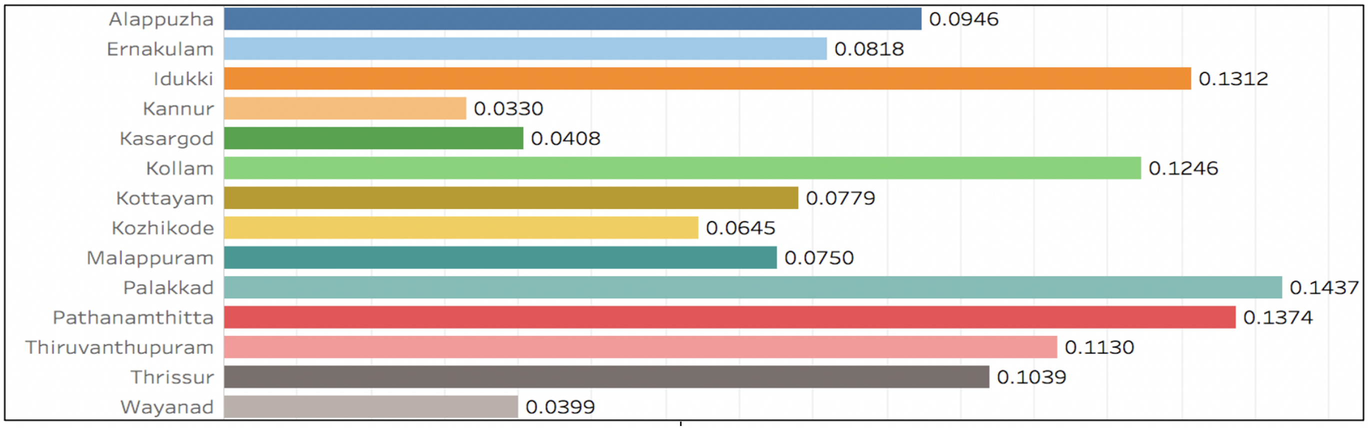


Figure 7.2: Scheduled Tribe Population by District (Proportion)

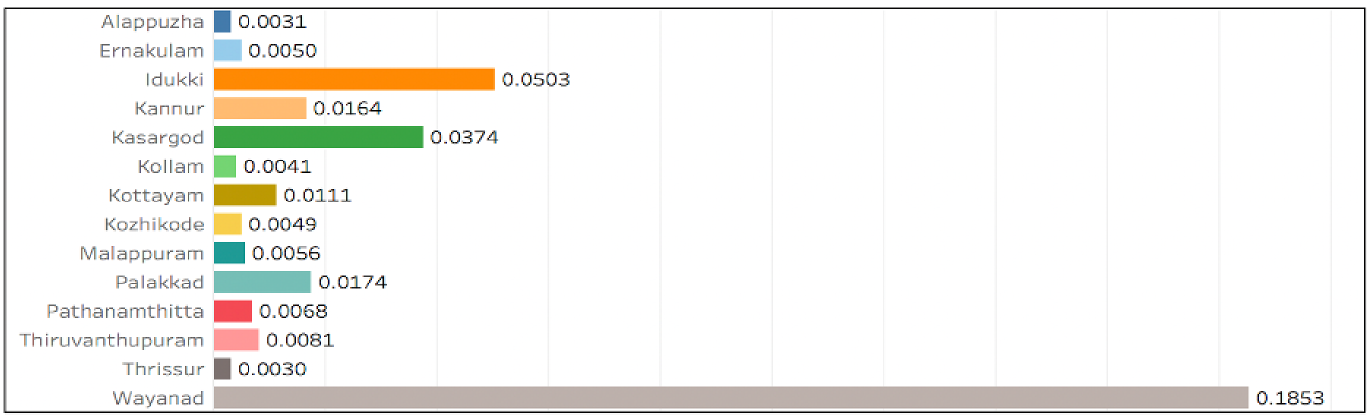
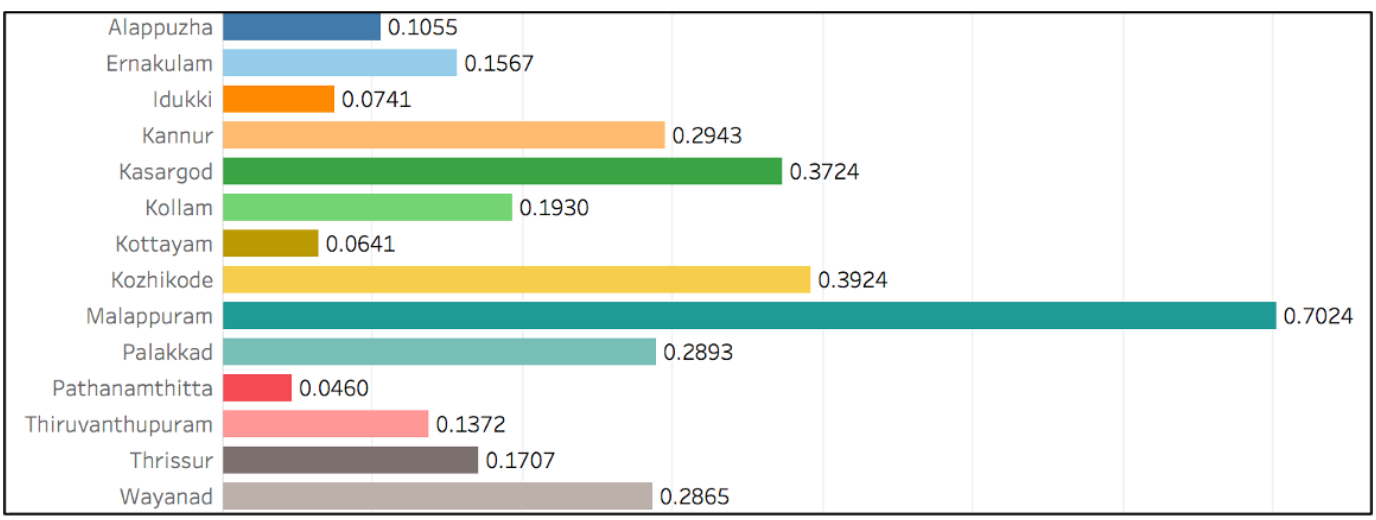


Figure 7.3: Muslim Population by District (Proportion)



There are also vast regional differences in tribal representation. For example, in Wayanad, 18.5% of the population is classified as Scheduled Tribe, compared to only .31% in Alappuzha ("India 2011 Census" 2011). The heterogeneity across districts provides strong variation to exploit in regression analysis (see above Figures).[4]

In addition to its demographics, Kerala has a rich availability of data on education and industry, statistics which cannot be easily found in other Indian states due to weaker infrastructure and poor data storage. Kerala also publishes examination results online, cross-checks national government school-wise data, and supplies extensive socioeconomic data at the district level. This provides a range of interesting covariates, including the inputs, pedagogy, and enrollment shares of the schools, which can be linked to the test results data for cross-school analysis.

7.2 Reviewing the Central Hypotheses of Interest

Recall, the central null hypotheses of interest are defined as:

H_1 = *There is no relationship between* ***access*** *and* ***learning outcomes.***

H_2 = *There is no relationship between* ***inputs*** *and* ***learning outcomes.***

H_3 = *There is no relationship between* ***school type*** *and* ***learning outcomes.***

The primary central hypothesis (H_1) is derived from the theoretical underpinnings of conflicts over the purpose and potential of Indian education as a tool for elite domination on the one hand and a platform for liberation on the other hand, as well as from the limited literature engaging directly with links between access and learning outcomes. The null hypothesis suggests there is no relationship between access to educational institutions or student resources and performance within these institutions. In other words, schools with more disadvantaged students enrolled perform comparably to schools with more "elite" students enrolled.

[4]Figures 7.1, 7.2, and 7.3 were constructed using data from the 2011 Indian Census.

The first theory of elite domination suggests a negative relationship between access and equity. That is, the recent inclusion of disadvantaged students in public schooling should reflect itself in lower test scores for schools with higher enrollments of the disadvantaged, and in the disadvantaged castes themselves compared to the elite both because of negative spillovers in public education and positive spillovers for the elite in private schooling. The second theory of a platform for liberation suggests either a positive or null relationship between access and performance. That is, the inclusion of disadvantaged students will either have no relationship to performance or a positive relationship between greater access and stronger performance. In the setting of Kerala, a leader in educational outcomes and inclusion, I expect a null relationship between learning outcomes and caste enrollment.

The next hypothesis (H_2) is derived from the micro-strand of literature concerned with the relationship between inputs into education (i.e. funding, infrastructure, technological provision, teacher qualifications or training) and learning outcomes. Does an increase in the inputs into education have a measurable impact on student learning outcomes? Academic literature suggests this is generally not the case, though bundles of certain inputs (for example, the quality of teachers and their access to classroom resources) might reveal a positive relationship.

The final hypothesis (H_3) relates to the recent debate over school choice, and the significance of the private provision of education in the presence of low-performing government schools. The literature does not come to formative conclusions about whether or not private schools truly outperform public schools, nor whether low-fee private schools are truly accessible to the poor. The null hypothesis suggests there are no differences between private and public schools in the dimensions of learning outcomes. That is, there are relatively similar shares of disadvantaged groups' enrollment in both types of school, and students in these groups perform comparably regardless of school type.

Chapter 8

Reviewing the Data and Constructing Key Variables

8.1 SSLC Test Scores

The data used to test my central hypotheses include a combination of reports, databases, and surveys collected from both the central and state government. To gather comprehensive test scores at the school level, I scraped standardized examination results from Kerala's Secondary School Leaving Certificate (SSLC) for the Summer of 2019 (last year available). Students take the SSLC at the end of the 10th Standard, which is the final year of secondary school. The examination is based on the curriculum of the Kerala State Board of Education and the Keralan Department of General and Higher Education and is administered by the Keralan Board of Public Examinations.[1,2,3]

These entities fall under the jurisdiction of the Ministry of Secondary Education. Students are graded on their performance in the following subjects: Language (I and II)[4]; English; Hindi; Social Science; Physics; Chemistry; Biology; Mathematics; and IT. Depending on their overall scores, students are either classified as "EHS" (eligible for higher study) or "NHS" (not eligible for higher study).

The SSLC constitutes one of the most significant examinations in a student's life, as

[1]For more information on state, national, and private curricula streams, see Appendix A.
[2]The SSLC scores were collected from the Keralan government.
[3]You can find additional data here here.
[4]Language I refers to "first language" while Language II refers to "second language". Different schools in Kerala conduct instruction in either English (more common in private schools) or Malayalam (more common in public schools), which constitutes the student's first language.

the score an individual receives determines their eligibility and placement in higher study. Thus, the SSLC becomes a precondition for further educational attainment. After being awarded a certificate conditional on passing and a score for their performance, students compete for placement in higher secondary education (from the 11th - 12th Standard) or may instead choose to attend a vocational training institute (Trines 2019).

Results for the SSLC are provided for both individuals (through a seven-digit registry number) and schools (through a five-digit school code). To preserve the anonymity of individual students taking the examination, I aggregated all data to the school-level, by scraping data from Kerala's Department of Education, which provides information on school codes for every registered school in the state.[5] I then cross-checked this information with the Department of Higher Secondary Education's Integrated Examination Management System portal (iExaMS), which provides school codes at the revenue (district) and education (sub-district) levels for all schools registered for the 2019 SSLC examinations.[6]

Using a Python web-scraper, I automated the process of data collection at the school level by entering school codes into the SSLC results webpage.[7] Results were then compiled in PDF format across every district and sub-district in Kerala, for all registered schools which took the 2019 examination. In total, I gathered results for 3,041 schools and 437,482 students. After cleaning and correcting for measurement error, a total of 2,702 schools were analyzed.

Students from all secondary schools which follow the state curricula are required to sit the exams, so there is no reason to expect selection bias on any given exam date.The district-wise proportion of schools represented in the data is as follows: Alappuzha (6.8%); Ernakulam (11.7%); Idukki (5.2%); Kannur (7.5%); Kasaragod (5.2%); Kollam (8.3%); Kottayam (5.3%); Kozhikode (6.6%); Malappuram (10.4%);

[5]The SSLC scores were collected from Kerala's state website.
[6]Data on school codes were cross-checked from this website.
[7]See Appendix C for an overview and scripts detailing how the data was scraped using Python.

Palakkad (6.8%); Pathanamthitta (6.3%); Thiruvanthapuram (9.0%); Thrissur (9.5%); and Wayanad (1.5%).

The scores collected were converted from PDF format to Excel, using Tabula software.[8] After cleaning the converted data for missing results, the scores were converted from letter grades to numeric.[9] These numeric scores were then aggregated to the school-level from individual results.

Figure 8.1: Average Scores by District and Proportion of Un-Aided Schools

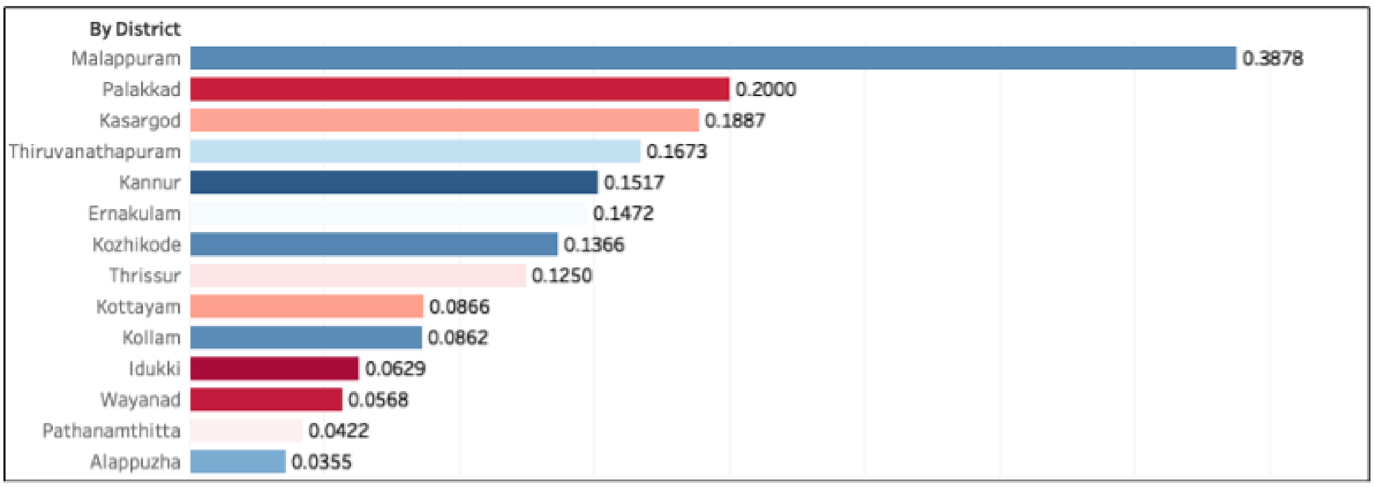
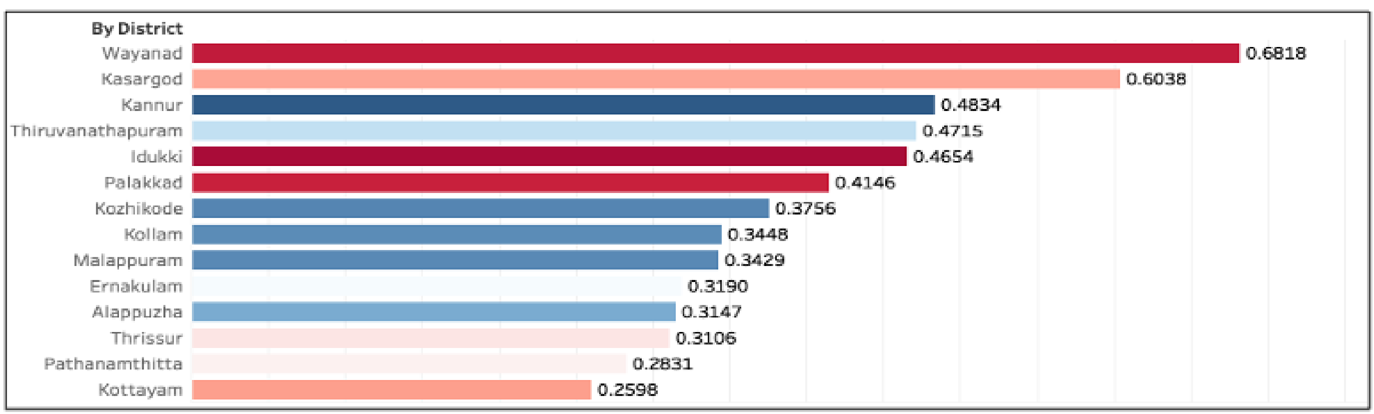


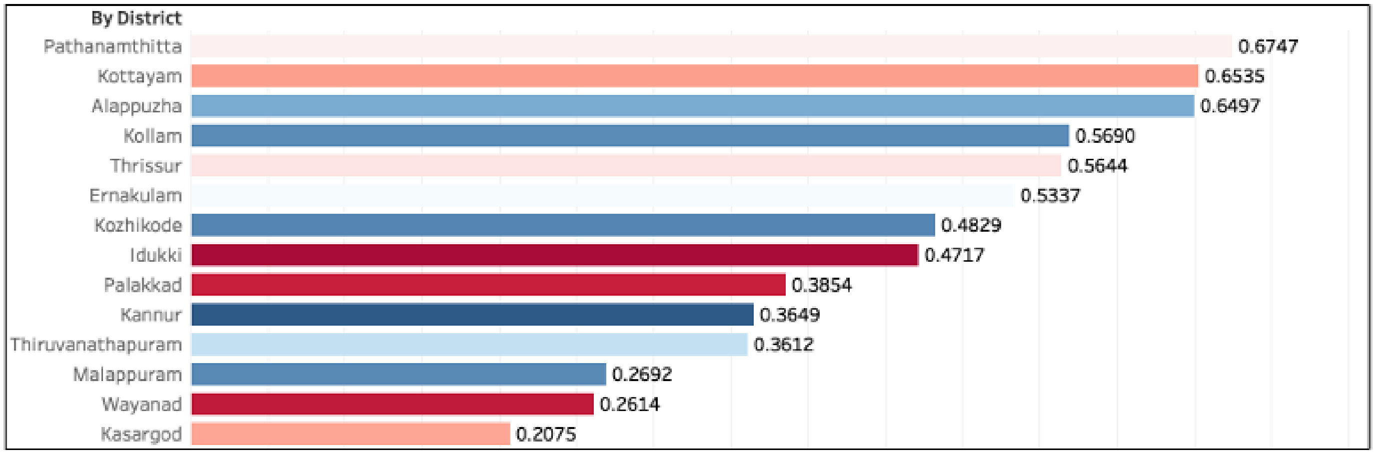
Figure 8.2: Average Scores by District and Proportion of Government Schools



[8]See Appendix C for information on the Tabula conversions.

[9]The conversions into numeric are as follows: 9 = A+; 8 = A; 7 = B+; 6 = B; 5 = C+; 4 = C; 3 = D+; 2 = D; F = 1. These scores were then multiplied by 10 (this improves variance of the scores). Thus, the lowest possible score is a 10, and the highest is a 90. This may seem unintuitive; but these codings were based on Kerala's own scoring system. More information on this conversion is provided in Section C.

Figure 8.3: Average Scores by District and Proportion of Aided Schools

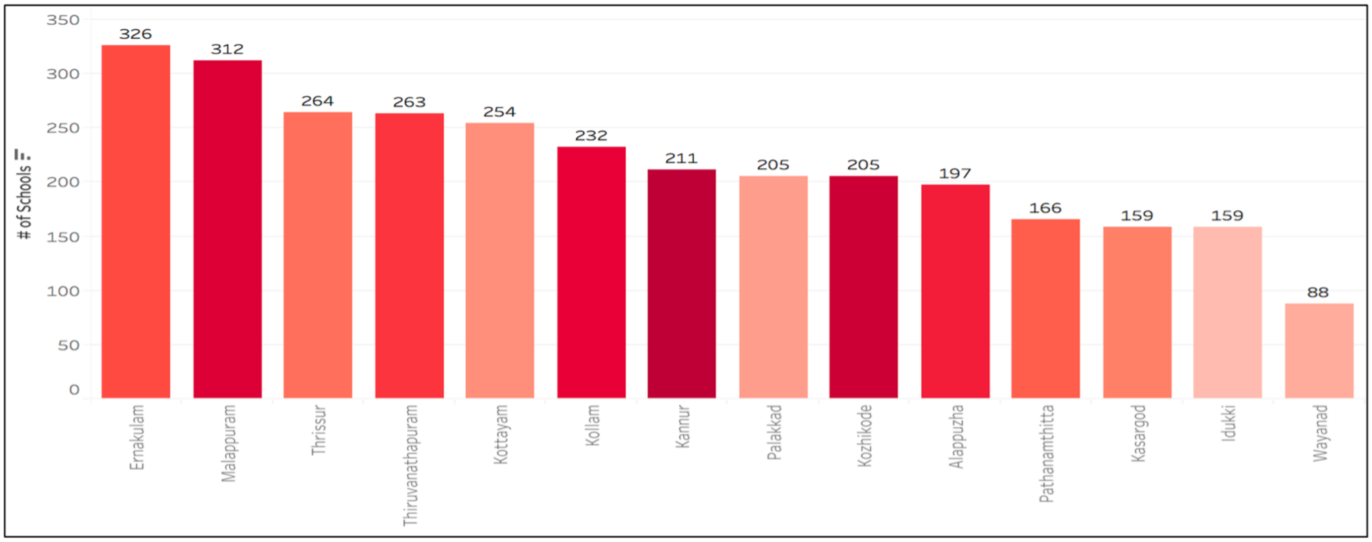


Above are selected preliminary statistics on mean SSLC test scores by district and school type. The color gradient represents mean scores (red is a poor score, white is average, and blue is strong), while the numbers appended to each bar represent the relative proportions of the respective school types in each district.

The district of Kannur appears to have the highest test scores regardless of school type (the bar in the darkest shade of blue). This despite having nearly half of its enrollments in government schools, and only 15% in un-aided private schools. Note that Kannur has the lowest proportion of Scheduled Caste individuals, and near to the lowest for Scheduled Tribe Individuals (see Figures 1 and 2). Meanwhile, Wayanad consistently performs worst across school categories. This is interesting considering the large proportion of ST individuals in the district relative to the rest of Kerala. Lastly, Pathanamthitta is dominated by a large proportion of government schools, which might allow for an isolation of the school type effect (after district-level controls).

Figure 7 examines average scores of schools (by district) conditional on their school size. The darker the shade of red, the higher the scores. Wayanad once more provides an interesting case. Not only does it have the worst test scores, it also has the fewest schools.

Figure 8.4: Average Scores by District and Number of Schools



8.2 UDISE School Report Cards

The iExaMS portal, which was used to cross-check school codes, also provides two additional labels for each school: the type of school (i.e. Aided, Un-Aided, or Government), as well as the UDISE code of the school. UDISE, or the Unified District Information System on Education, constitutes my second major data source. In 1994, the District Primary Education Programme (DPEP) was launched by the central government as a precursor to Sarva Shiksa Abhiyan ("UDISE+ Booklet" 2019).

The program approached educational change at the district-level, in an attempt to tailor policies to local conditions and involve panchayat communities in the decision-making process. In conjunction with DPEP, the government established an Education Management Information System (EMIS) for data collection and storage, which gave rise to UDISE. This national system of data collection has since expanded to cover the entirety of India, down to the district level, for all levels of schooling. In 2015, UDISE was designated the sole government platform for data on Indian schooling (Bordoloi and Kapoor 2018).

In addition to their various reports and policy recommendations, UDISE also generates detailed school report cards for every registered school in India, in collaboration with the National Institute of Educational Planning and Administration. Each report card provides rich data at the school-level, mapped to the UDISE code of each school. To collect this data, I ran another web-scraper through Python[10] for the 2016-2017 academic year, matching the UDISE codes of the schools with 2019 SSLC examination data to the school report card website.[11]

Roughly 270 of the schools with SSLC scores (less than 10%) did not have U-DISE School Report Cards for the 2016-2017 academic year and were thus not included in analysis. The missing observations were well-balanced across districts. 2,702 schools were analyzed in the 2016-2017 academic year, which represents roughly 57% of the entire universe of secondary schools in Kerala.[12] By merging the school report card data with my examination results, I construct a rich cross-sectional dataset. I also link the coordinates (latitude and longitude) of the schools to spatial data at the neighborhood level, using data from the Ministry of School Education and Literacy.[13]

UDISE data informs both policymakers and academics, so the accuracy of the data has been closely scrutinized. UDISE follows a rigorous system of checks for quality of their Data Capture Format (DCF). The entire process of collection takes eight months (from July to February) and involves a wide range of actors. Preliminary entry of school-level data is assigned to the headmaster of the school. These headmasters are trained in data entry by UDISE staff. If the school does not have the resources to conduct data entry (i.e. no electricity or internet), a UDISE staff member inputs the data directly in collaboration with the headmaster.

The staff include District and Block Coordinators, Block Resource Center Coordinators (BRCs) and Cluster Resource Center Coordinators (CRCs). CRCs are positioned within the largest of a cluster of 10-15 schools and are responsible for monthly

[10] More information on the web scraper is provided in Section C of the Appendix. A sample school report card is also provided as a screenshot.
[11] School Report Cards were collected from here.
[12] According to the government's School Geographic Information System platform.
[13] Data was collected from here.

checkups within this cluster. They report to the BRCs, who then report to the District and Block Coordinators. Data collection is cross-checked in the field at the 100% level (at each cluster) by the CRC Coordinators and at the 25% level (at each block) by the BRC Coordinators. An overall review of the data at the district level is conducted in early November, followed by consistency checking and cross-checks at the sub-district level in December. Prior to any sampling, UDISE hires an independent third-party to perform a 5% random sampling check (Bordoloi and Kapoor 2018). The range of actors and various quality checks of UDISE are neatly illustrated by UNESCO on the following page.

Figure 8.5: Range of Actors in UDISE Data Collection (Bordoloi and Kapoor, 2018)

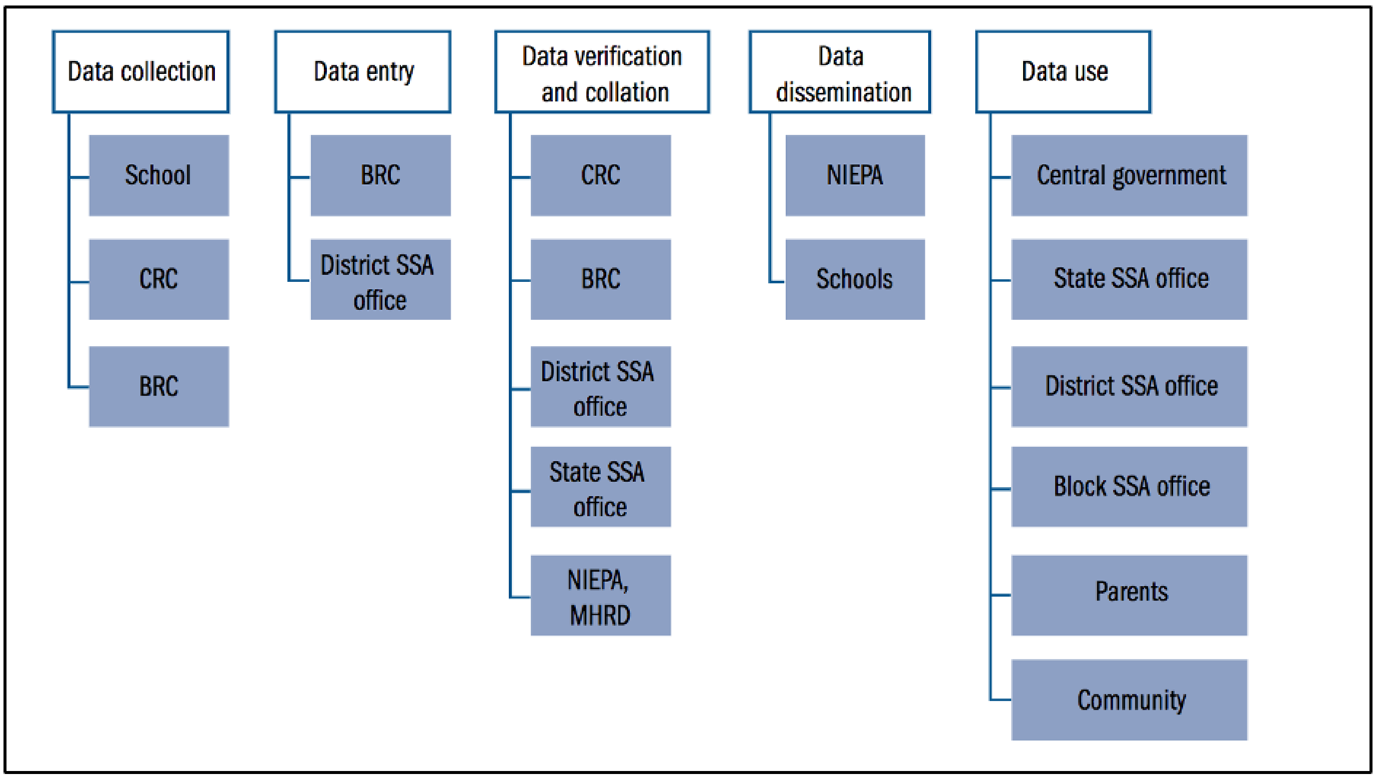
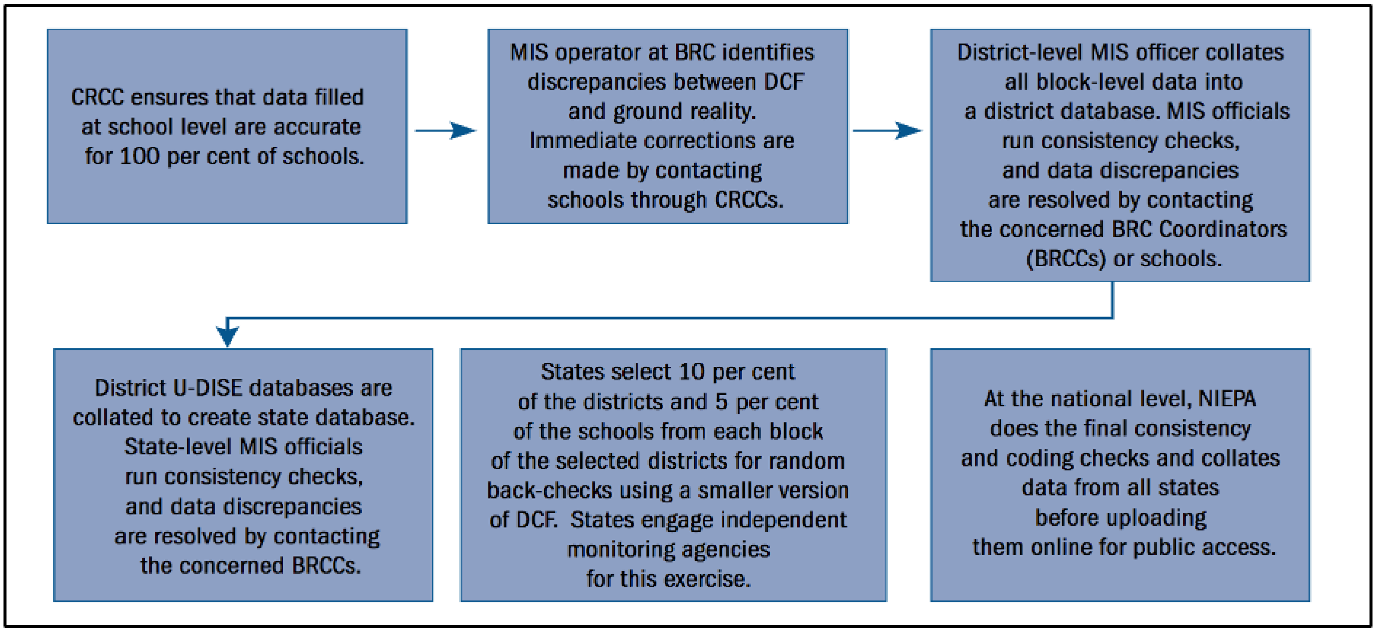


Figure 8.6: Processes of UDISE Data Collection (Bordoloi and Kapoor, 2018)



Despite the rigorous process of data collection, UDISE data is imperfect and may be subject to measurement error. To begin, because the process of collection takes an extended period of time, UDISE issues data with a lag of one year (e.g. AY 2018-2019 is issued in 2020). The uploading and cross-checking of the data inputted by headmasters is also dependent on transfers of information between District and Block MIS officials, CRC Coordinators, and BRC Coordinators, often without a clear audit trail. Sometimes, the mandated cross-checks are not conducted due to time or cost restraints ("UDISE+ Booklet" 2019).

In order to improve the processes of data collection and the rigor of quality checks, UDISE recently established a new platform, UDISE+. This system will improve on some of the inconsistencies in UDISE data but is not yet available for public access. In order to correct for potential measurement error in the standard UDISE report cards, I create a series of checks for key variables (enrollment, inputs, examination results), which recalculate key variables using the report cards to confirm calculated average and percentage values.

In addition to the school-level features, I also collect a series of relevant controls to include as additional covariates in the regression analysis. These include district-level demographic and socioeconomic indicators collected from the Keralan Department of Economics and Statistics,[14] 2011 Census Data, and detailed information on education from the UDISE+ dashboard.[15] Additional controls include urban economics data on the number of universities; proportion of literate adults or those with college degrees; health indicators for families and children; industry data and factory statistics; and any sub-district level data scraped through reports from the Keralan Department of Economics and Statistics. A review of the main dependent and independent variables of interest are presented below.

[14] Data collected from various reports here.
[15] Additional data collected from here.

8.3 Construction of Key Variables

8.3.1 Dependent Variable

1. **2019 SSLC Test Scores**: Average score per school across subjects. Data is available for 2,702 schools across the state of Kerala (constituting roughly 60% the universe of secondary schools in the state). The classification of letter grades into numeric is as follows: F = 10; D = 20; D+ = 30; C = 40; C+ = 50: B = 60; B+ = 70; A = 80; A+ = 90. Scores range from a low of 47.4 to a high of 89.6.

8.3.2 Independent Variables

1. **Enrollment of Disadvantaged Groups**: Enrollment counts of Girls, Scheduled Caste (SC), Scheduled Tribe (ST), Other Backward Caste (OBC), and Muslim students. Counts were also converted to proportions (shares), for both Class X students (those who take the SSLC) and overall school-wise statistics (for robustness checks) in a secondary analysis.

2. **Inputs**: Input covariates include the teacher-pupil ratio, student-classroom ratio, binary variables for presence of a playground, presence of a science lab, and the number of computers functional in a school. Funding information (i.e., grants received from the government) was also analyzed.

3. **School Type**: Categorical variable coded as one of three school types: government; aided; and un-aided (private). Most schools in the sample are aided or government schools.

4. **Pedagogy**: Teacher information (count of disadvantaged teachers; count of teachers with professional qualifications; count of regular and contract

teachers) was collected, as were indicators relating to the classroom environment, including the student-classroom ratio and student-teacher ratio.

5. **Additional Covariates**: Additional covariates from the school report card include (1) geographic information (sub-district and village name; location) and (2) school information (gender type; enrollments for Class X and the entire school). From official government reports, the covariates include (3) industry information (salaries; number of factories) and (4) demographic information (population; per capita income; number of universities; health indicators), at the district-level. Additional geospatial data was also collected for analysis.

Chapter 9

Descriptive Analysis

Before proceeding with empirical analysis in a regression framework, I present initial trends in the data. First, it is interesting to note the distribution of disadvantaged groups across school profiles (and the prevalence of these profiles in the sample). I focus on three features: school type; location; and gender type. Following this, I examine scores across these profiles. Next, I present district-level estimates of enrollments across OBC, SC, ST, and Muslim enrollments. Note that these social identities are not mutually exclusive. Lastly, I present the initial correlations linking my covariates of interest to average scores.

Based on the literature which posits elite self-selection bias, we might expect to see high percentages of caste enrollment in aided and government schools, as opposed to un-aided (private) schools. The sample demographics align more with the literature which promotes private schooling as an affordable, low-fee alternative to government schooling. Nearly half of the students at un-aided schools are Muslim, and more than two-thirds come from Other Backwards Castes. Meanwhile, the enrollment shares of these groups in aided and government schools is far lower. Note that while this general trend holds for OBC and Muslim students (who account for a majority/sizeable minority of students in Kerala respectively), they do not apply to Scheduled Caste and Scheduled Tribe students. These groups are almost exclusively enrolled in government and aided schools. In total, 49% of schools are aided; 38% are government; and 13% are private.

Figure 9.1 (on the following page) examines disadvantaged enrollments by core demographics. Kerala is a rural state, and thus most schools are categorized as rural (76%). Enrollment shares do not appear to vary significantly by location; and average test scores of urban and rural schools are nearly identical (70.57 in rural schools vs. 70.96 in urban schools).

The variation of gender type is not pronounced in the data; almost all the schools in the sample are coed. Only 6.3% of schools are girls only, and 2.5% are boys only. The relative shares are similar across gender types; though there is a noticeable jump in ST enrollments for boys only schools. This is an interesting result which I do not expand on here but might suggest that lower castes shift expenditures towards sons rather than daughters (much of the literature supports this).

Figure 9.1: Disadvantaged Enrollments by Core Demographics

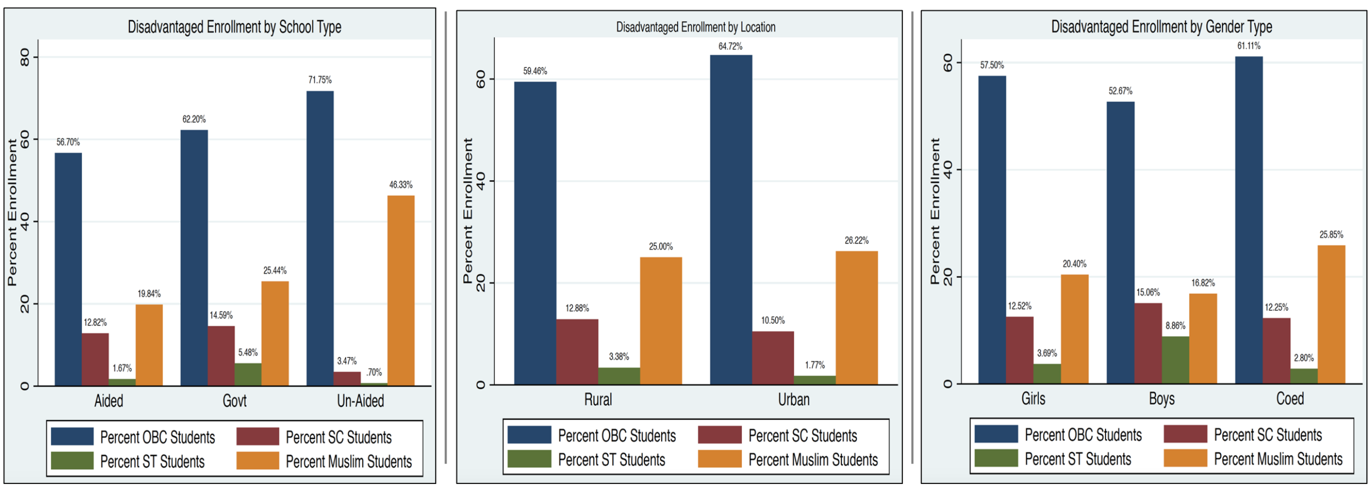


Figure 9.2 (on the following page) presents a histogram of the average scores in the sample. There is a relatively wide spread of scores. As expected, we see a rough bell curve approximating a normal distribution.

Figure 9.2: Distribution of Mean SSLC Scores

Next, I turn to another dimension of analysis, which links average scores to different school features. The variation in scores by location is not pronounced (see above), so I restrict descriptive analysis to school type and gender type, where the differences are more robust. In Figure 9.3, the first graph aligns with previous findings by the literature. There is a strong private school effect; the average private school performs nearly 10 points higher on the SSLC than aided or government schools, which share similar outcomes. In addition to the unique characteristics of these school types, the observed differences might also be explained by self-selection (either based on idiosyncratic talent, which sorts the most able into private education, or based on resources and endowments).

Figure 9.3: Scores by School Profile

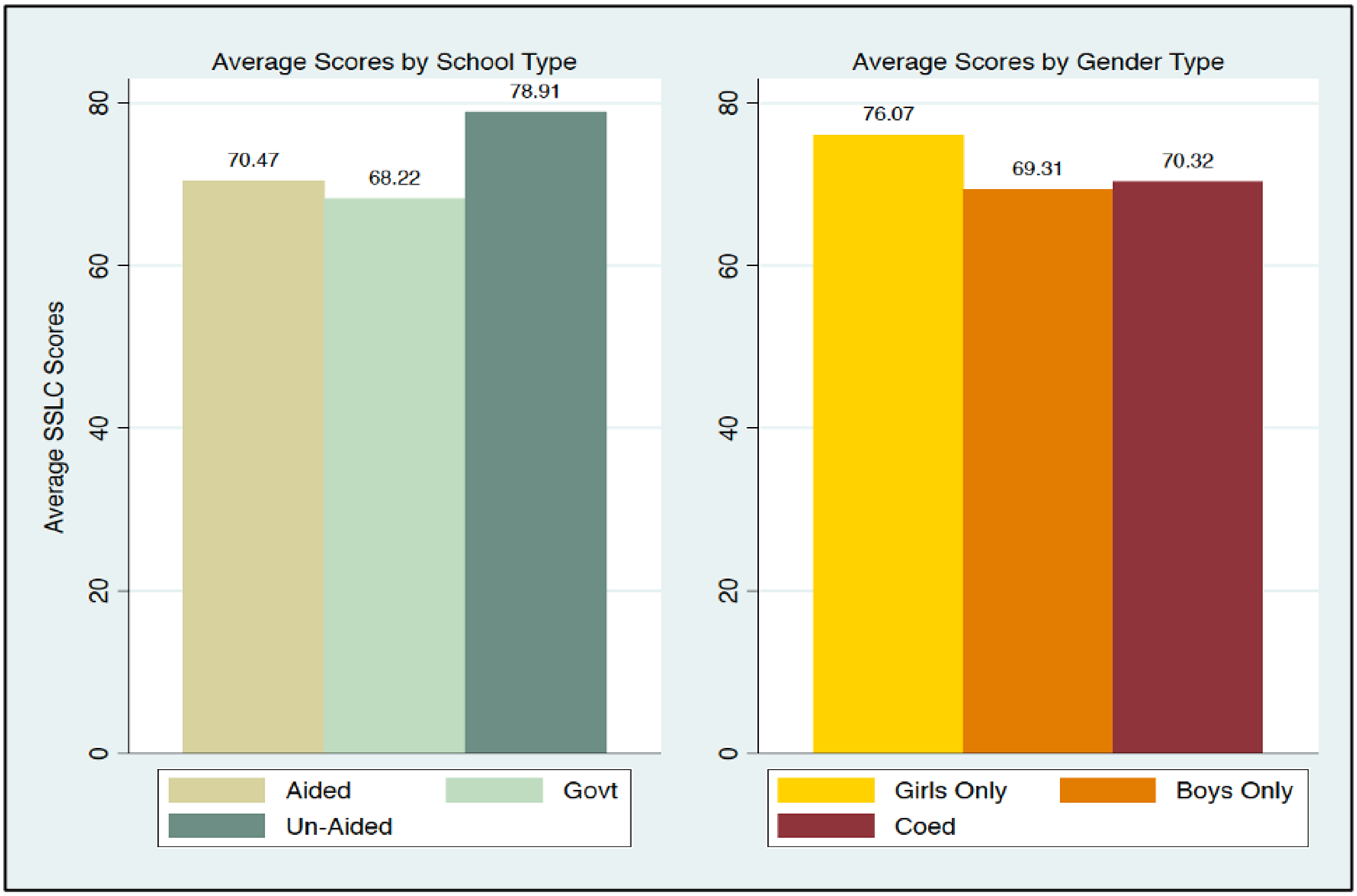


Figure 9.4 examines the variance (or lack thereof) across covariates of interest for teacher pedagogy. It's evident that certain "school profiles" have stronger variation in the sample than others. At government schools (the first bar), we can see a large degree of variation in Qualified teachers and OBC teachers. At the bottom of the graph, we see a small sliver which represents the variation in ST teachers. The variation in SC teachers at this school type is so small, that it's not visible in the first bar. This trend continues across both Aided and Private schools. While there is strong variation in OBC and Qualified teachers, the variation in the other

indicators is not substantial. In fact, ST and SC teachers are relatively rare in the sample. Across school profiles, an average school has 17 OBC teachers, but only 2 SC teachers. Most schools have only a single ST teacher.[1] The lack of variation in these indicators should be considered when interpreting the regression analysis in the following section.

Figure 9.4: Variance Across Indicators for Teacher Features (by School Type)

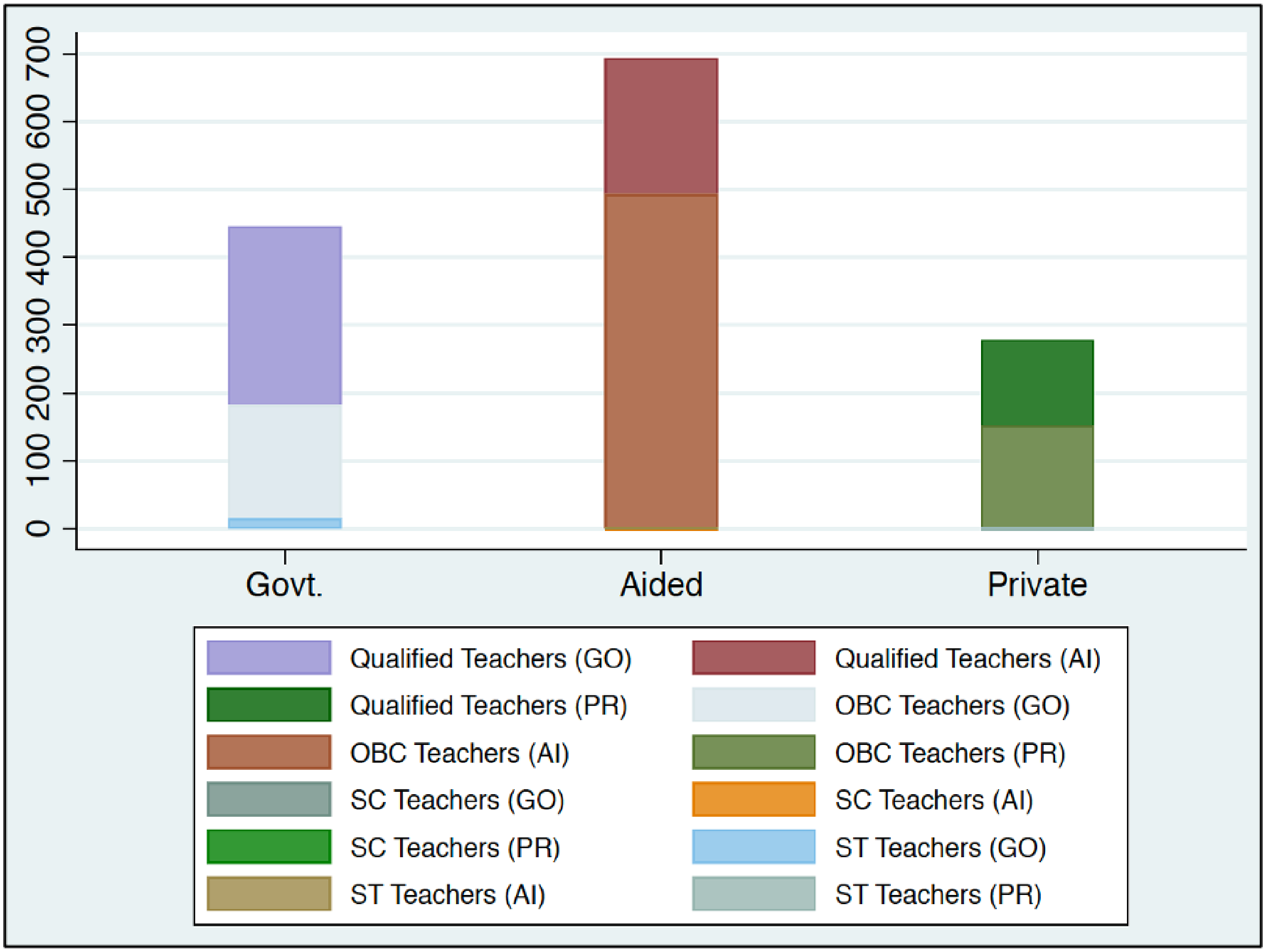


Figure 9.5 presents scatterplots for the key relationships of interest, linking disadvantaged enrollments to average SSLC scores.[2] The scatterplot for OBC enrollment suggests a slightly positive relationship with scores, while SC and ST have slightly negative relationships. There appears to be no relationship between Muslim en-

[1]Unfortunately, the data does not highlight the religious composition of the teacher base.
[2]Note: two outliers have been removed.

rollments and scores. There are a few outliers in the data (which constitute large government schools).[3] It also appears that OBC and Muslim enrollments are far higher than SC or ST enrollments, which aligns with state demographics.

Figure 9.5: Scatterplot Matrix of School Scores Against Disadvantaged Enrollments

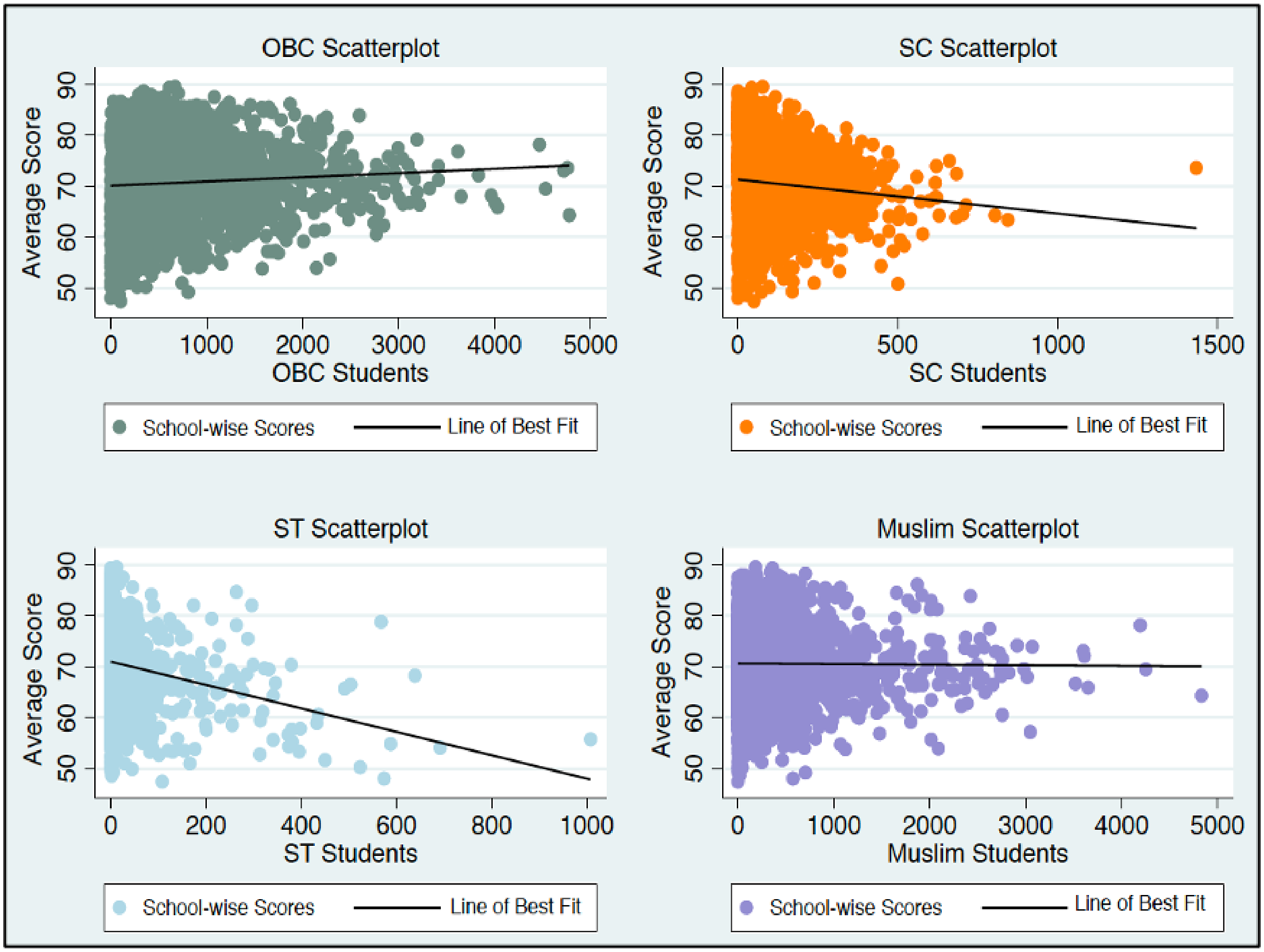


Figure 9.6 plots a correlation heatmap across the indicators of interest. To read the figure, start at the bottom left corner ("Average Scores") and move up the matrix vertically. The colors relate to the strength of the initial correlations of interest. The heatmap suggests the null hypotheses are likely to hold. That is, there does not seem to be a strong relationship (positive or negative) between disadvantaged enrollments

[3]The proceeding regression analysis was run both with and without outliers, with no significant differences across the results. I choose to include the estimations including the outliers in this paper, to better approximate the true universe of secondary schools in Kerala.

and school wise test scores. Student expenditures seem to be slightly positively correlated with scores, suggesting that schools in districts where households spend more on education are likely to perform better. We also observe the large positive effect described in the literature and note that female enrollment weakly links to higher scores. The number of Muslims enrolled, and the number of qualified teachers also appear to have a positive correlation with one another. Both the matrix and the heatmap illustrate the overlap in student identities. A large number of OBC students identify as Muslim. Overall, the covariates of interest do not appear to substantially link with scores. Note that these descriptive figures do not account for controls and interactions, so we cannot draw conclusive evidence.

Figure 9.6: Correlation Heatmap of Key Indicators

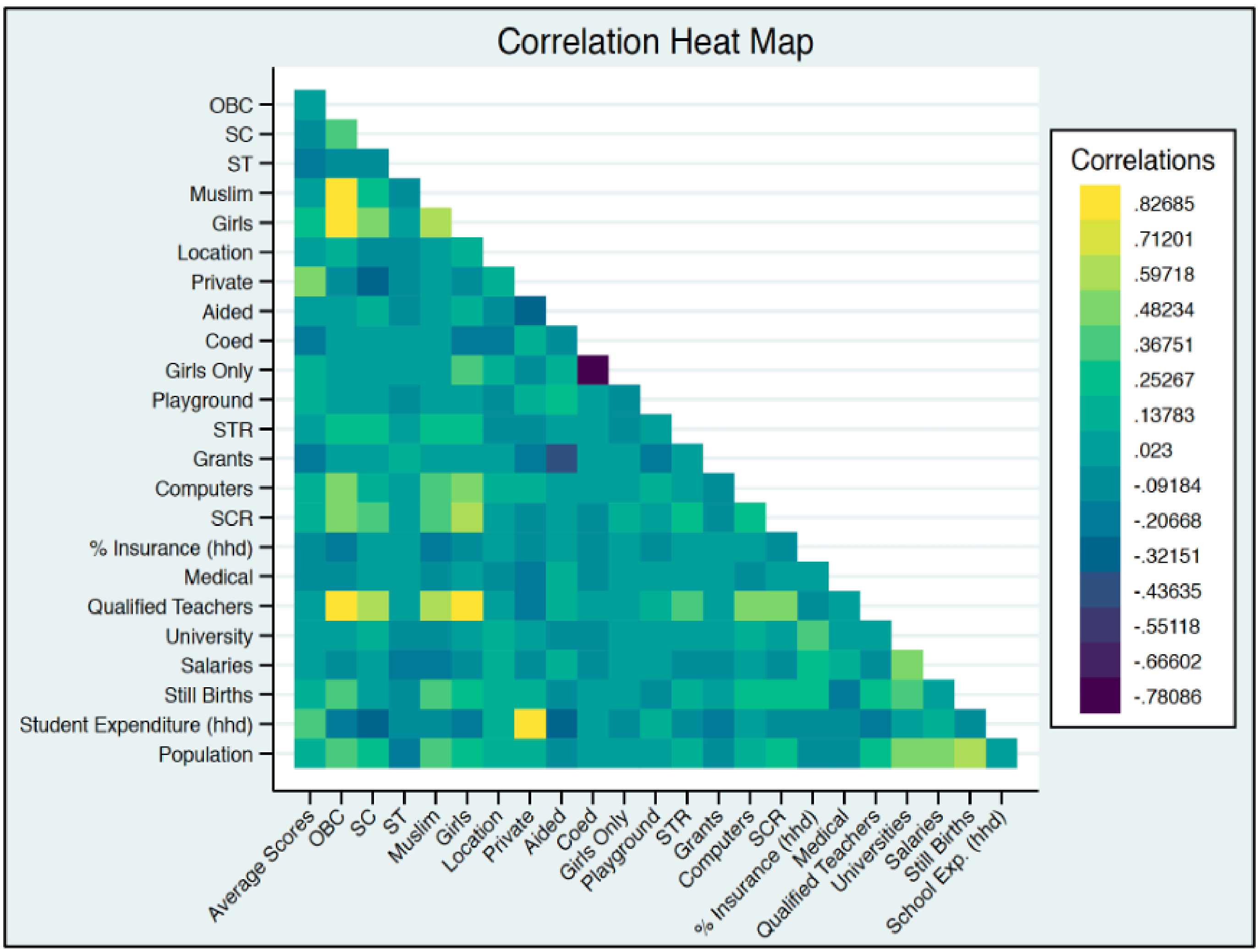


Figure 9.7: Disadvantaged Enrollments by District

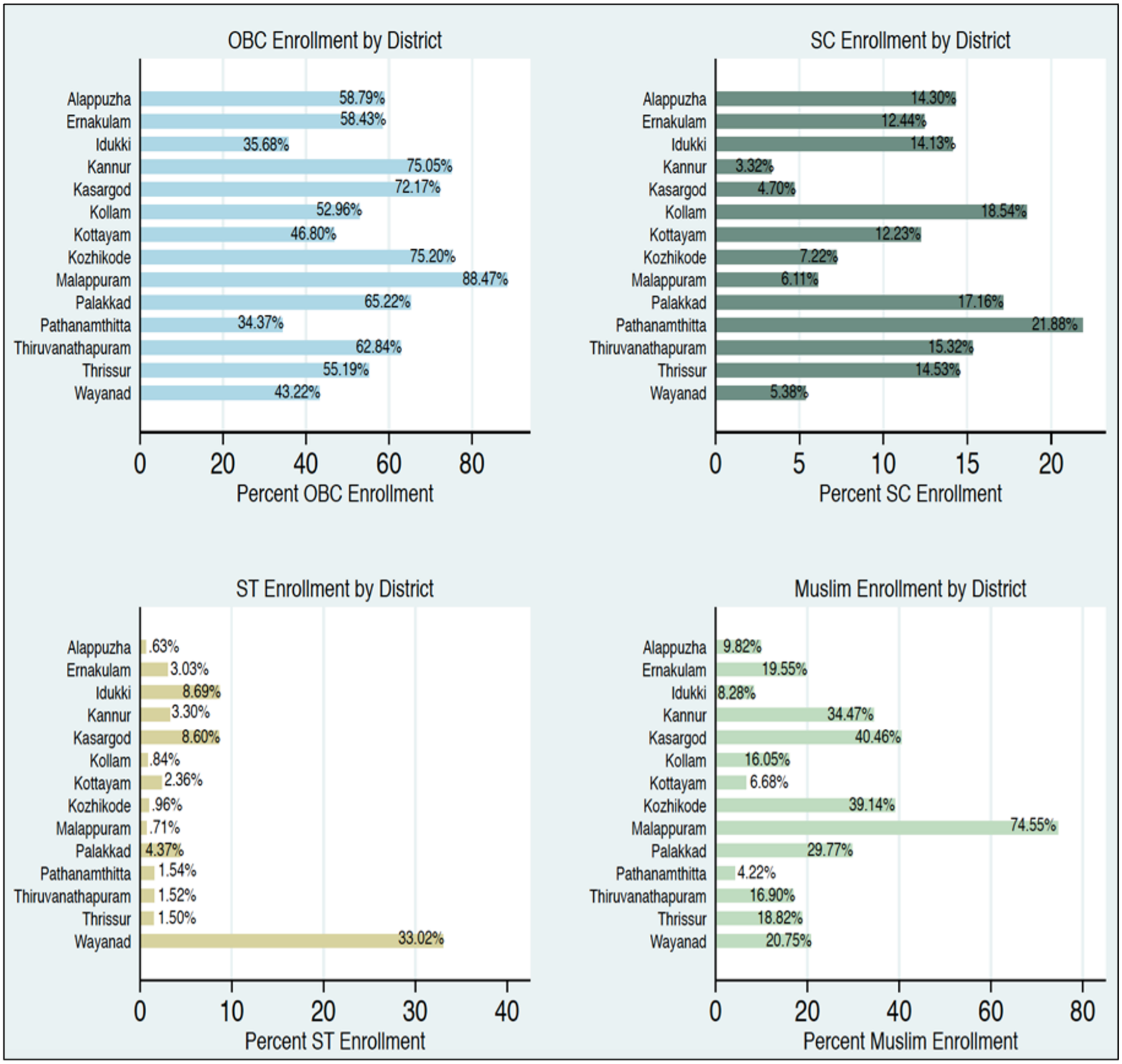


Figure 9.7 (above) provides a brief overview of disadvantaged enrollments aggregated to the district level. In many of the districts, OBC students constitute the majority of enrollments. SC enrollments are also sizeable across many districts (including Pathanamthitta and Palakkad), while ST enrollments is small in all districts except Wayanad. Muslims constitute a large majority in Malappuram.

Figure 9.8 (following page) presents descriptive statistics for the covariates of interest. Note that district-level covariates have been italicized. The dependent variable, average SSLC scores, is normally distributed. Enrollment shares across schools are quite volatile; reflected in the skewness of the distributions and large variation of the indicators. Many schools record enrollments at or near zero for ST and SC individuals.[4] This is expected, as it is very unlikely that the distribution of disadvantaged students across districts and schools in Kerala is perfectly normal. Furthermore, these groups (particularly ST children) constitute a minority of the population in most districts.

I also note a negligible gender gap in enrollment. Both girls and boys both enroll in schooling; and the previous figures suggest schools with higher enrollments of girls perform better on average. This suggests that families are willing to invest in the education of all their children, rather than boys alone.[5] Another important finding is the lack of disadvantaged teachers across school types. On average, schools only hire between one and two teachers from Scheduled Caste or Scheduled Tribe backgrounds. The skewness for Scheduled Tribe teachers is also incredibly high (most schools do not hire any Scheduled Tribe teachers).

Lastly, it is interesting to briefly explore the input indices of schools in the sample. The majority of schools have playgrounds, science labs, or medical check-ups but there is some variation in those schools which lack inputs. The average grants received (for public schools in the year 2016-2017) was 24,705 rupees. However, large government schools received larger grant allocations than smaller aided schools.

[4] 132 schools have an SC enrollment of 0; and 1,212 schools have an ST enrollment of 0.

[5] Whether or not this is the case for disadvantaged groups (i.e. Scheduled Tribe) needs to be explored further.

Figure 9.8: Key Descriptive Statistics

Variables	Obs.	Mean	Std. Dev.	Min	Max	p1	p99	Skew.
SSLC Scores	2702	70.664	7.214	47.448	89.579	53.344	85.689	-.179
Total OBC	2702	643.46	659.89	0	7244	0	2971	2.327
Total SC	2702	102.56	109.327	0	1433	0	482	2.453
Total ST	2702	16.695	58.106	0	1008	0	299	7.03
Total Muslim	2702	311.50	525.259	0	6478	0	2495	3.391
% OBC	2702	60.679	25.047	0	100	0	100	-.244
% SC	2702	12.331	10.476	0	90.686	0	50.731	1.694
% ST	2702	3.007	10.981	0	100	0	65.188	5.947
% Muslim	2702	25.283	29.279	0	159.31	0	100	1.231
Total Enrollment	2702	972.58	789.093	15	10152	51	3450	2.139
Percent Girls	2702	48.677	18.056	0	100	0	100	1.045
Girls Class X	2702	76.653	89.998	0	993	0	417	2.614
SC Class X	2702	16.394	20.152	0	236	0	94	2.904
% SC Class X	2702	12.939	12.651	0	100	0	57.143	1.743
ST Class X	2702	2.376	8.863	0	193	0	42	8.457
% ST Class X	2702	2.966	11.164	0	100	0	64	5.835
OBC Class X	2702	105.90	138.7	0	2117	0	659	3.647
% OBC Class X	2702	60.787	26.483	0	100	0	100	-.291
% Muslim Class X	2702	25.962	30.953	0	122.353	0	100	1.135
SCR	2702	33.177	12.145	1.51	134.286	6.333	61.95	.409
STR	2702	86.712	104.115	1.909	1698.5	7.975	516.667	5.945
SC Teachers	2702	1.666	2.612	0	21	0	11	2.366
ST Teachers	2702	.605	2.569	0	62	0	7	15.83
OBC Teachers	2702	16.708	18.273	0	164	0	80	1.802
Qualified Teachers	2702	39.035	23.898	1	278	7	111	1.583
Grants	2702	24705	33057.7	0	528500	0	103500	5.259
Computers	2702	13.882	12.074	0	200	0	55	4.587
Hhd. Expenditure (rupees)	2702	106.54	8.645	89	121	89	121	-.233
School Expenditure (rupees)	2702	7117.8	3938.745	2920	23097	2920	22924	2.294
% Hhd. with Computers	2702	30.94	9.568	11.4	54.1	11.4	54.1	.002
Major Universities	2702	4.976	4.525	1	17	1	17	1.624
Factories	2702	534.79	337.389	59	1156	59	1156	.814
Factory Output (rupees)	2702	117000	1990000	54971	6701510	54971	6701510	2.364
Value Added (rupees)	2702	4430000	1550000	1133809	6904548	1133809	6904548	-.339
School Count	2702	376.88	105.032	129	579	129	579	.25
Classroom Count	2702	3951.65	1338.125	1123	6494	1123	6494	.064
Secondary Enrollment	2702	84638.7	36833	28713	170926	28713	170926	.791
SC Enrollment	2702	9.545	3.599	3	15.4	3	15.4	-.143
ST Enrollment	2702	1.515	2.662	.3	20.2	.3	20.2	5.134
OBC Enrollment	2702	60.755	15.866	30	87.4	30	87.4	-.19
Muslim Enrollment	2702	27.386	19.884	4.9	73.8	4.9	73.8	1.13
Live Births	2702	30007.2	17070.24	8547	72499	8547	72499	1.239
Still Births	2702	118.655	89.734	23	348	23	348	1.014
Akshayas	2702	236.245	59.397	84	331	84	331	-.307
% Women Attended School	2702	95.611	2.493	88.79	98.74	88.79	98.74	-1.29
% Hhd. Health Insurance	2702	47.037	7.849	37.25	61.79	37.25	61.79	.19
% Stunted Children	2702	18.827	4.608	12.41	27.65	12.41	27.65	.203
% Anemic Women	2702	33.249	7.336	22.42	42.88	22.42	42.88	-.186
Value Added (rupees)	2702	4430000	1550000	1133809	6904548	1133809	6904548	-.339
Per Capita Income (rupees)	2702	164000	24981.15	121388	204050	121388	204050	-.098
Population	2702	2740000	954000	838395	4408415	838395	4408415	-.223

Chapter 10

Empirical Model and Results

Based on the literature, my derived relationships of interest, and the descriptive analysis above, I construct a cross-sectional regression, in a simple OLS estimation framework. To improve robustness, I also conduct a secondary fixed effects analysis and include an alternative specification using principal component analysis (PCA) to correct for minor collinearity in the other two specifications. In each case, I generate three regression models to estimate my central relationship of interest, between disadvantaged access and learning outcomes. Each model focuses on one of three different elements of minority status and social identity in India, all incredibly important in the context of schooling: the caste of a student, the religion of a student, and the gender of a student (male or female). These models are run independently to correct for collinearity between the enrollment rates of different disadvantaged groups, which overlap.[1] In order to incorporate different functional relationships of interest, I include interaction terms in the primary analysis. In the secondary analysis, I estimate these relationships using both counts and proportions, at the level of the entire school and the level of Class X.

[1]For example, a single student could be a female Muslim categorized as "Scheduled Tribe".

10.1 General Regression Framework

I specify my primary specification of interest as:[2]

$$\textbf{(1)}\ \textit{St. Average SSLC Scores}_i = \alpha_0 \ + \ \beta_i \ \textit{St. Disadvantaged Enrollments} \ + \ \Sigma\,\delta_i\,x_i \ + \ \Sigma\,\lambda_i\,x_i + \ \Sigma\,\omega_i\,x_i \ + \ \Sigma\,\gamma_i\,x_i \ + \ \Sigma\,\mu_i\,x_i \ + \ \Sigma\,\kappa_i\,x_i \ + \ \Sigma\,\eta_i\,x_i \ + \ \Sigma\,\theta_i\,x_i \ + \ \Sigma\,\zeta_i\,x_i \ + \epsilon_i$$

Where "St. Disadvantaged Enrollments" in Equation **(1)**[3] refers to the *standardized* count of a given disadvantaged group (by Caste, Religion, or Gender) enrolled. "δ_i" is a vector of covariates for school features, including location, private, aided, coed, girls only. "Location" is a binary which refers to whether a school is located in an urban or rural area. "Private" and "Aided" refer to school type (government schools are the excluded category). "Coed" and "Girls Only" refer to gender type (boys only schools are the excluded category and are rare in the sample.

"λ_i" represents a vector of input characteristics, including playground, medical check-up, computers, and grants. "Playground", "Science Lab", and "Medical Check-Up" are binaries which indicate whether a school has a playground, a science lab or conducts medical check-ups; "St. Computers" refers to the standardized count of computers at the school; and "St. Grants" refers to the standardized amount of government grants received (note that Un-Aided schools do not receive grants).

"ω_i" is a vector of covariates linking to school pedagogy, including STR, SCR, St.

[2]See Appendix D for the full regression specification. Note that this was not the initial specification; the model was re-specified, and covariates were adjusted to account for collinearity and measurement error in the preliminary results.

[3]Legend for Regression Equation: β Coefficients – Enrollment Shares
δ Coefficients – School Features
λ Coefficients – School Inputs
ω Coefficients – School Pedagogy
γ Coefficients – Interactions of School Features
μ Coefficients – Interactions of School Inputs and Pedagogy
κ Coefficients – Interactions of School Features and Pedagogy
η Coefficients – Interactions of Enrollments and Pedagogy
θ Coefficients – District Controls
ζ Coefficients – Interactions of District Controls and School-Level Covariates

Qualified, St. SC Teachers, St. OBC Teachers, St. ST Teachers. "STR" refers to the school's student-teacher ratio; and "SCR" refers to the student-classroom ratio. "St. Qualified" refers to the standardized count of qualified teachers; "St. OBC Teachers" refers to the standardized counts of OBC teachers; "St. SC Teachers" refers to the standardized counts of SC Teachers; and "St. ST Teachers" refers to the standardized counts of ST Teachers.

"γ_i" is a vector of district-level socioeconomic controls, including household access to computers; household expenditure; school expenditure; major universities; salaries; population; and percentage of households with access to health insurance.

Lastly, "μ_i" "κ_i" "η_i" "θ_i" and "ζ_i" represent a series of interaction terms between the main covariates of interest, and "ϵ_i" is the unobservable error term. Because the units for most variables are standardized, the general interpretation of any coefficient is as follows: a shift in "x" standard deviations from the mean value of a specific independent variable results in a shift of "y" standard deviations from the mean value of the dependent variable (average test scores).[4]

$$\textbf{(2)}\ \textit{St. Average SSLC Scores}_i = \alpha_0 + \beta_i\ \textit{St. Disadvantaged Enrollments} + \Sigma\,\delta_i\,x_i + \Sigma\,\lambda_i\,x_i + \Sigma\,\omega_i\,x_i + \Sigma\,\gamma_i\,x_i + \epsilon_i$$

Equation **(2)**[5] incorporates a fixed effects regression for district-level features, which controls for the location of a school within a specific district through indicator variables (where the dummy takes a value of 1 when a school is in the given district).

[4]The following variables have been standardized: OBC Children Enrolled, SC Children Enrolled, ST Children Enrolled, Muslim Children Enrolled, Girls Enrolled, Functional Computers, Total Grants Received, Total Contract Teachers, Total Qualified Teachers, Total SC Teachers, Total ST Teachers, Total OBC Teachers.

[5]Legend for Regression Equation: β Coefficients – Enrollment Shares
δ Coefficients – School Features
λ Coefficients – School Inputs
ω Coefficients – School Pedagogy
θ Coefficients – District Controls

The district dummy variables are collinear with the district level covariates (and thus cannot be included in the fixed effects analysis) and with certain school features in the primary model, so I construct an alternative model (PCA – Principal Component Analysis) which collapses the district covariates into three columns. In the secondary estimation, the range of covariates slightly expands. For ease of interpretation, the interaction terms were removed from these additional specifications. I extend analysis to include the counts and proportion of the disadvantaged groups enrolled in either the entire school or Class X alone.

10.2 Estimation Results for the Multivariate OLS Specification

In this section, I discuss results from the primary models of interest. All variables were standardized when possible (counts of disadvantaged groups, grants received, number of contract teachers, number of qualified teachers), and standard errors were clustered by school to eliminate potential heteroskedasticity. Because the OLS specification incorporates interaction terms (binary x binary; continuous x binary; and continuous x continuous), we must interpret the coefficients with caution. Before exploring the marginal effects of the interactions themselves, I begin by analyzing individual covariates of interest in Table 10.1 (below).

Table 10.1: Primary OLS Specification

VARIABLES	(1) Caste	(2) Religion	(3) Gender
Total OBC Enrolled	-0.0393		
	(-0.041)		
Total SC Enrolled	-0.132***		
	(-0.0228)		
Total ST Enrolled	-0.151***		
	(-0.0235)		

Table 10.1 continued from previous page

VARIABLES	(1) Caste	(2) Religion	(3) Gender
Total Muslim Enrolled		-0.151***	
		(-0.0308)	
Muslim School		-0.397***	
		(-0.114)	
Total Girls Enrolled		0.220***	
		(-0.0319)	
Urban vs. Rural	-0.471***	-0.397***	-0.394***
	(-0.0704)	(-0.0708)	(-0.0706)
Private School	0.829***	0.804***	0.845***
	(-0.185)	(-0.212)	(-0.179)
Aided School	-0.0407	-0.0138	0.0105
	(-0.0644)	(-0.0646)	(-0.0642)
Coed School	-0.0696	-0.0358	-0.237**
	(-0.118)	(-0.116)	(-0.119)
Girls Only School	1.040***	1.027***	0.579***
	(-0.164)	(-0.154)	(-0.171)
Qualified Teachers	0.378***	0.284***	0.112**
	(-0.0501)	(-0.0436)	(-0.0494)
Total OBC Teachers	-0.198***	-0.0994***	-0.125***
	(-0.0266)	(-0.0216)	(-0.0207)
Total SC Teachers	-0.0860***	-0.112***	-0.109***
	(-0.0275)	(0.0265)	(-0.0265)
Total ST Teachers	-0.114***	-0.117*	-0.132***
	(-0.0347)	(-0.014)	(-0.0127)
Functional Computers	0.0302	0.0090	0.0456
	(-0.0567)	(-0.0569)	(-0.0566)

Table 10.1 continued from previous page

VARIABLES	(1) Caste	(2) Religion	(3) Gender
Playground	0.0876**	0.0884**	0.0844**
	(-0.0345)	(-0.0349)	(0.0351)
Science Lab	-0.170**	-0.192***	-0.147**
	(0.0667)	(-0.0668)	(-0.0653)
Medical Facility	-0.129	-0.338	-0.136
	(-0.268)	(-0.263)	(-0.27)
Total Grants Received	0.0971	0.0606	0.0705
	(-0.0686)	(-0.0703)	(-0.071)
Student-Teacher Ratio	-0.0002	-0.0002	-0.0003
	(-0.0002)	(-0.0002)	(-0.0002)
Student-Classroom Ratio	0.0134***	0.0125***	0.00818***
	(-0.00177)	(-0.00168)	(-0.00175)
Urban Private School	0.324**	0.214	0.245*
	(-0.132)	(-0.133)	(-0.132)
Urban Aided School	0.309***	0.248***	0.243***
	(-0.0878)	(-0.0875)	(-0.0874)
Private Girls Only School	-0.858*	-0.630*	-0.806*
	(-0.463)	(-0.381)	(-0.449)
Aided Girls Only School	-0.342***	-0.225*	-0.314**
	(-0.128)	(-0.12)	(-0.126)
Student-Teacher Ratio * Grants	-0.0004**	-0.0003*	-0.0003*
	(-0.0002)	(-0.0002)	(-0.0002)
Student-Teacher Ratio * Computers	0.0001	8.84E-05	0.0001
	(-0.0001)	(-0.0001)	(0.0001)
Student-Classroom Ratio * Grants	-0.0003	0.0004	-6.92E-05
	(-0.0021)	(-0.0021)	(-0.0022)

Table 10.1 continued from previous page

VARIABLES	(1) Caste	(2) Religion	(3) Gender
Student-Classroom Ratio * Computers	0.0055***	0.0059***	0.0042**
	(-0.0018)	(-0.0017)	(-0.0017)
Qualified * Computers	-0.023***	-0.013*	-0.016*
	(-0.0088)	(-0.0071)	(-0.0066)
Qualified * Grants	0.0147	0.0020	0.0034
	(-0.0214)	(-0.0194)	(-0.021)
Qualified * Household Computers	0.0229	0.0731***	0.066***
	(-0.0193)	(-0.0176)	(-0.0178)
Qualified * Household Expenditure	0.117***	0.0420**	0.093***
	(-0.0203)	(-0.0203)	(-0.0178)
Qualified * Private	0.0949	0.166**	0.163**
	(-0.0845)	(-0.0811)	(-0.0817)
Qualified * Aided	-0.165***	-0.131***	-0.127***
	(-0.0478)	(-0.0444)	(-0.0456)
OBC Teachers * OBC Students	0.0398***		
	(-0.0124)		
SC Teachers * SC Students	-0.0050		
	(-0.0096)		
ST Teachers * ST Students	0.0049		
	(-0.0038)		
Muslim School * Girls Only		-0.564**	
		(-0.271)	
Muslim School * Private		0.380**	
		(-0.158)	
Muslim School * Aided		-0.433***	
		(-0.144)	

Table 10.1 continued from previous page

VARIABLES	(1) Caste	(2) Religion	(3) Gender
Muslim School * Urban		0.16	
		(-0.136)	
% Household Health Insurance	0.0031	0.0009	0.0041
	(-0.0054)	(-0.005)	(-0.0054)
Household Computers	0.0118	0.0640***	0.0518***
	(-0.0208)	(-0.0195)	(-0.0194)
Household Expenditure	-0.104***	-0.102***	-0.073***
	(-0.0286)	(-0.0275)	(-0.0274)
Major Universities	0.00316	-0.0344	-0.0102
	(-0.0228)	(-0.0227)	(-0.0224)
Log of Salaries	0.0286	-0.0227	0.0258
	(-0.0333)	(-0.034)	(-0.0332)
School Expenditure	0.169***	0.235***	0.202***
	(-0.0585)	(-0.0622)	(-0.0575)
Log of Population	0.0504	0.262***	0.106
	(-0.0866)	(-0.0838)	(-0.0824)
*% Household Insurance * Medical*	0.0031	0.0068	0.0029
	(-0.0056)	(-0.0055)	(-0.0056)
Observations	2,702	2,702	2,702
R-Squared	0.369	0.360	0.357

Clustered standard errors in parentheses

*** p <0.01, ** p <0.05, * p <0.1

10.2.1 Primary Specification: Disadvantaged Enrollments

For schools with the mean number of SC teachers, an SC student enrollment one deviation above the mean links to a score .132 standard deviations below the average. For schools with the mean number of ST teachers, an ST student enrollment one deviation above the mean links to a score .151 standard deviations below the average. Schools with a Muslim enrollment one deviation above the mean score .151 deviations below the average. The result for OBC enrollments is also slightly negative, but insignificant. Interestingly, not all traditionally disadvantaged students are struggling. Schools with higher enrollment shares of girls as compared to boys seem to perform substantially better. Schools with female enrollments one deviation above the mean score .220 deviations above the average score. Private girls only schools perform roughly one deviation above the mean, and aided girls only schools score .657 deviations higher.

In Model **(2)**, I include additional interaction terms for disadvantaged enrollment, using the covariate "Muslim School" as a key element. Muslim School is coded as a binary, which takes on the value 1 if a school has a share of Muslim enrollment greater than 75% of the school's total (a conservative cutoff). Due to these conservative conditions, some of the school profiles captured in the interactions are relatively rare in the sample. For example, there are only 8 girls-only Muslim schools.[6] These Muslim girls-only schools score roughly .067 standard deviations above the mean. Meanwhile, Muslim private schools score .787 deviations above the mean and Muslim aided schools score .02 deviations above the mean. Thus, while the overall relationship between Muslim enrollments and scores is negative, at schools with a large majority of Muslim students (particularly at girls-only and private schools); the relationship becomes positive. This might suggest positive peer effects in Muslim communities, but also reveals the power of the private and girls only effects, which outweigh large enrollment shares of disadvantaged groups.

[6] There are 127 private Muslim schools; 81 aided Muslim schools; and 55 urban Muslim schools.

10.2.2 Primary Specification: Pedagogy and Inputs

Schools with the mean enrollment of OBC students, and a share of OBC teachers one deviation above the mean, score .198 standard deviations below the average. Schools with the mean enrollment of SC students, and a share of SC teachers one deviation above the mean, score .0860 standard deviations below the average. Lastly, schools with the mean enrollment of ST students, and a share of ST teachers one deviation above the mean, score .0049 standard deviations below the average. Schools with a playground tend to perform better than those without, across all three models. However, schools with a science lab tend to perform worse. The coefficients for medical checkups and computers at a school are insignificant.

Figure 10.1: Relationship between Grants Received and Average Scores

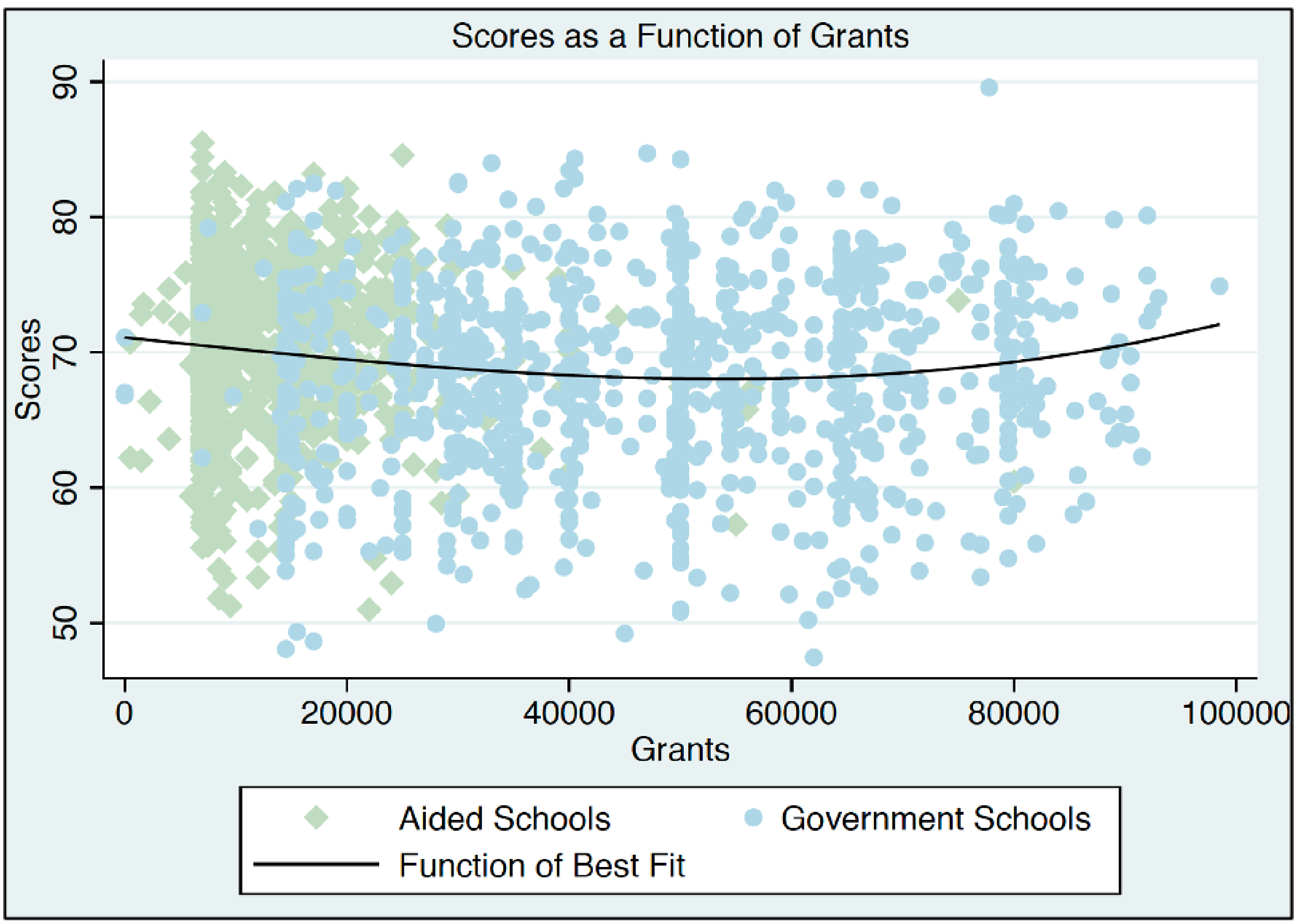


Figure 10.1 (above) illustrates the function of best fit describing the relationship between grants received and test scores (non-standardized) for both aided and government schools.[7] There appears to be no relationship between the amount of money a school receives and their test scores. The regression table, also, suggests an insignificant effect. Both results align with the literature on input-based education. It is interesting to note that a few aided schools received no grant funding for the 2016-2017 academic year, and most received small amounts of funding. Government schools were generally offered higher grant allocations. The district covariates were simply used as controls. However, a few coefficients are worth exploring further.

To begin, school expenditure has a significant, and positive relationship with scores. Districts with school expenditures (averaged across households) one deviation above the average perform between .17 and .23 standard deviations above the mean score. Thus, higher school expenditures link to higher scores, an intuitive result. Districts with larger populations also score better (according to Model 2), which might be linked to stronger resources and schooling systems in condensed, urban areas. The proportion of households in a district which computers also appears to influence scores. For example, in districts in which 10% of households own computers, schools with the average number of qualified teachers perform .64 standard deviations above the average score.

Table 10.2: Predictive Analysis by School Profile

School Type	***Location***	***Gender Type***
Aided: -0.2643	Urban: 0.4104	Coed: -0.0470
Government: -0.3387	Rural: -0.0124	Girls Only: 0.7498
Private: 1.1436		Boys Only: -0.1871

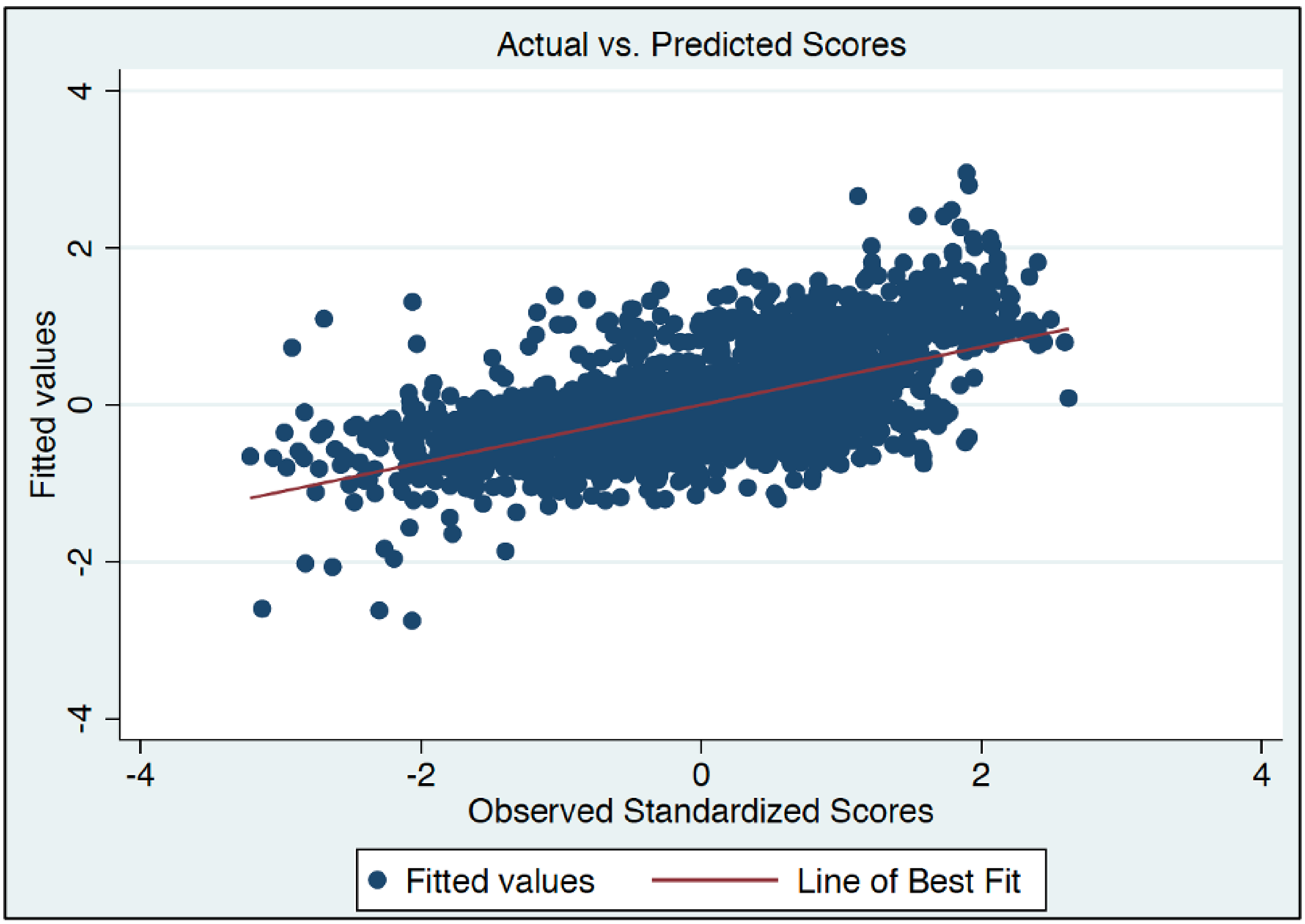
Above, I provide a table of predicted scores across school profiles. These predictions are measured in standardized units, just like the regression model. We can interpret the predictions as follows: a "typical" aided school will score .026 standard deviations

[7]Note that un-aided private schools don't receive government grants.

below the average test score of the sample. The predictive relationships are largest for girls only schools and private schools, both of which perform substantially better than the mean.

Figure 10.2 (next page) compares the fitted values from the predictions to the observed values of the test scores. There appears to be a relatively strong fit, suggesting the model was well-specified.

Figure 10.2: Goodness of Fit: Actual vs. Predicted Scores

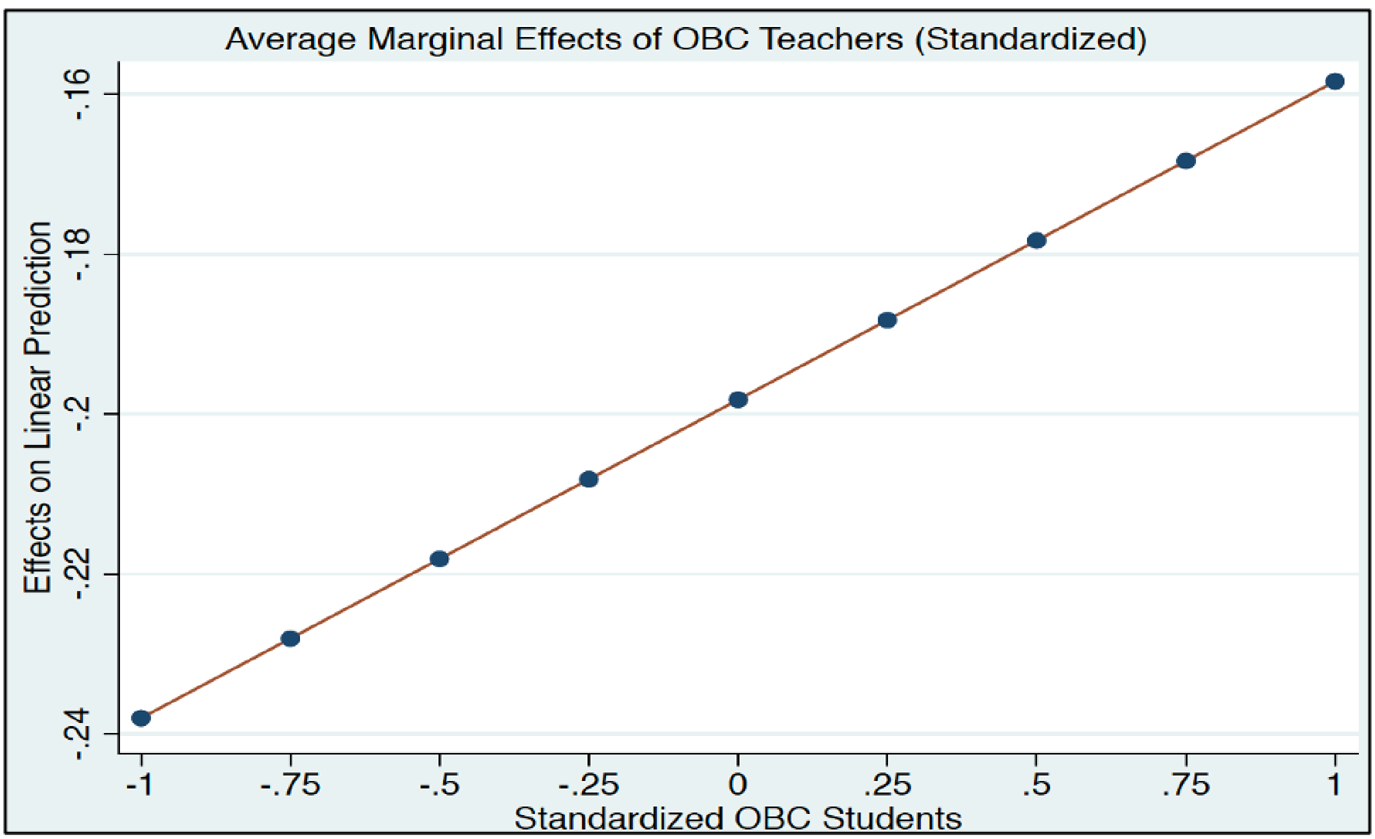


10.2.3 Primary Specification: Marginal Effects & Interaction Terms

Now, I turn to an analysis of the interaction terms. Each interaction term and their elements were tested for joint significance. Only those which were deemed both statistically and practically significant in Model **(1)** are analyzed below.

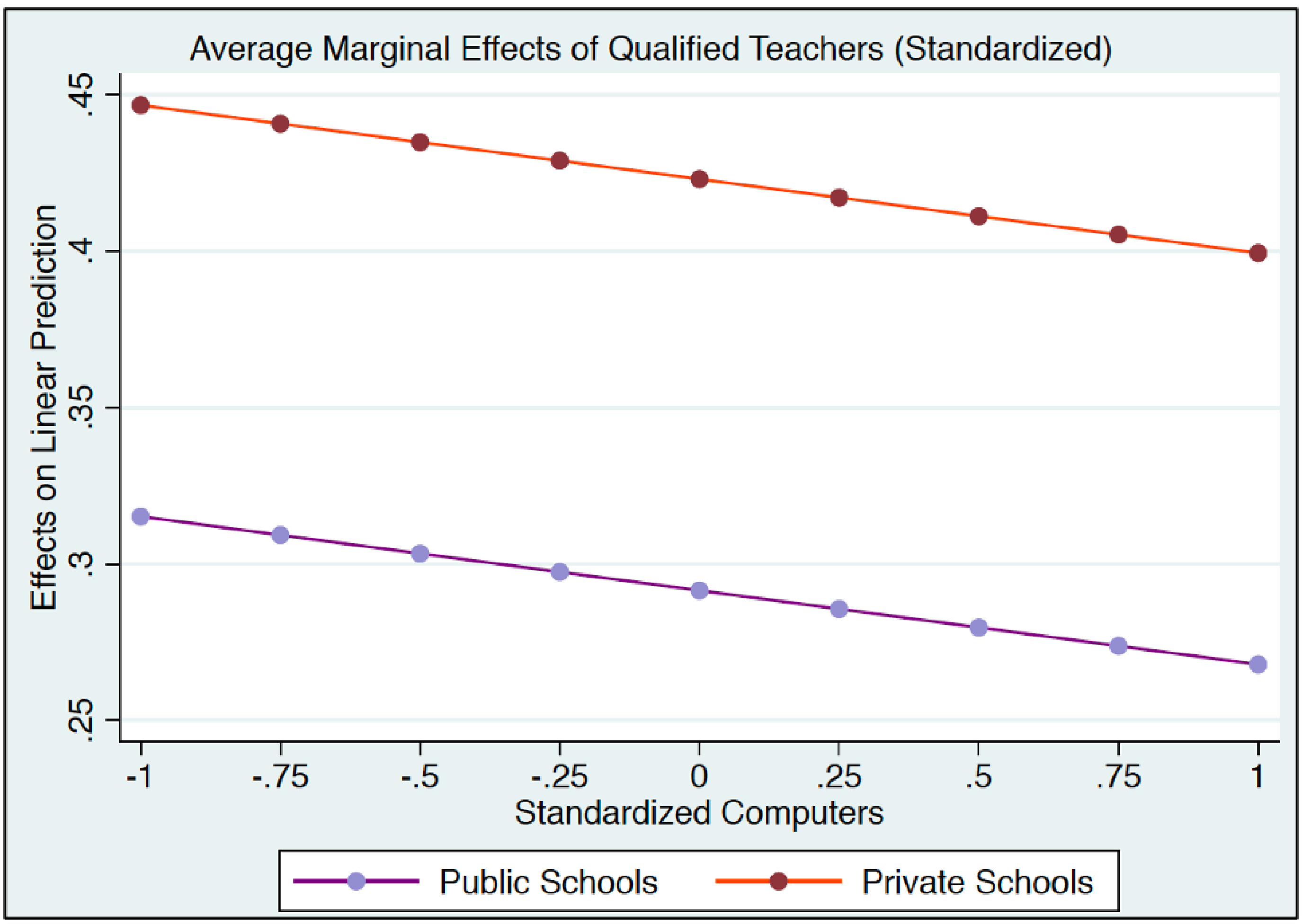
OBC Teachers x OBC Students: OBC Teachers have a negative marginal effect on test scores, however at schools with increasing numbers of OBC students, this negative effect rapidly approaches zero (diminishes). This might suggest positive co-ethnic bias, whereby OBC students are better able to identify and relate to fellow OBC mentors. We might expect these social similarities to bolster learning outcomes for this group of students. A similar result is found for ST Teachers x ST Students; however, the magnitudes are not practically significant.

Figure 10.3: Marginal Effects: OBC Teachers x OBC Students



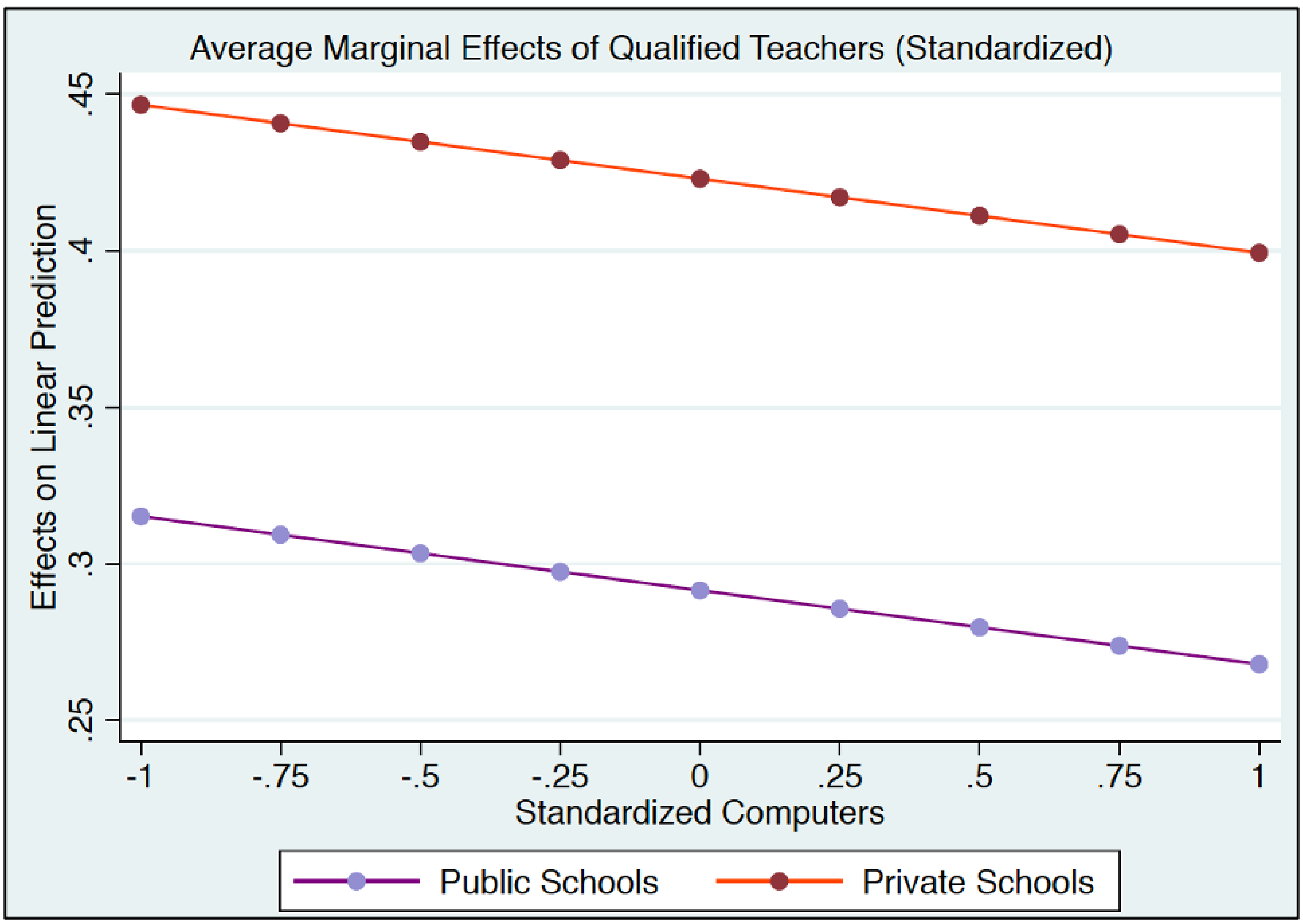
Qualified Teachers x Computers: Qualified teachers have a positive but diminishing marginal effect on scores as the number of computers at a school increase. In other words, at schools with a greater number of computers, the positive relationship between the number of qualified teachers at a school and that school's scores increases, but at a diminishing rate. This might be because of a substitution away from teachers, and towards technology-aided instruction at the extremes of the distribution. It is also interesting to note that computers have the same rate of returns across both private and public schools (despite potential differences in teacher quality). This provides illustrative evidence of the potential for an RCT comparing teacher quality across private and public teachers in Kerala (as we see parallel trends in the data).

Figure 10.4: Marginal Effects: Qualified Teachers x Computers



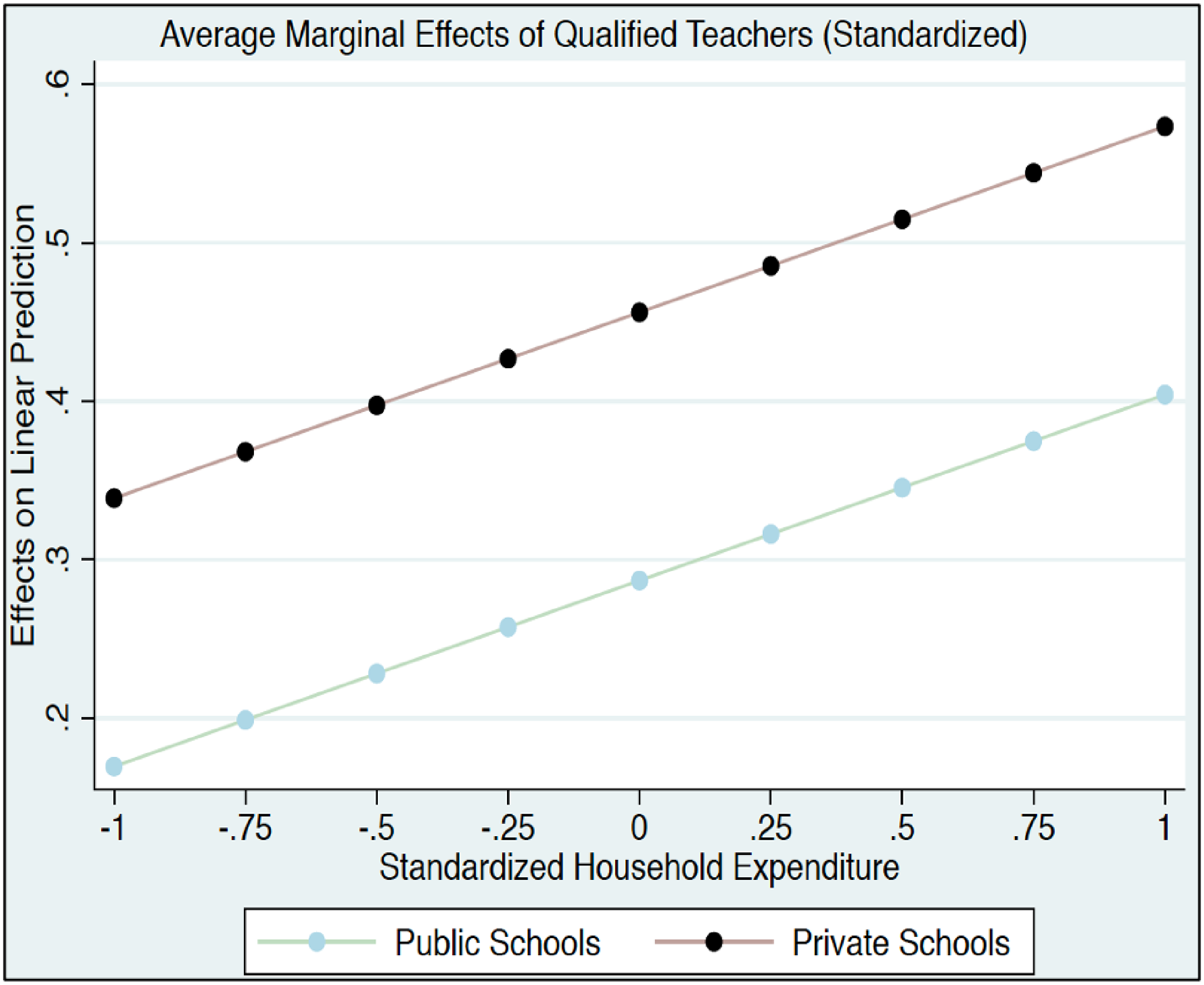
Qualified Teachers x Grants: Qualified teachers have a positive marginal effect as school grants increase. The change in the marginal effect is constant, and increasing, across various levels of grants. This interaction might be dominated by the qualified teacher effect. In other words, the grants themselves might not themselves contribute to higher scores (see Figure 22), but rather the number of qualified teachers at schools generate this positive relationship. For example, we might expect schools with higher funding (i.e. wealthier schools in wealthier districts) to have stronger teachers. Note that only public schools (government and aided) receive grants and are included.

Figure 10.5: Marginal Effects: Qualified Teachers x Grants



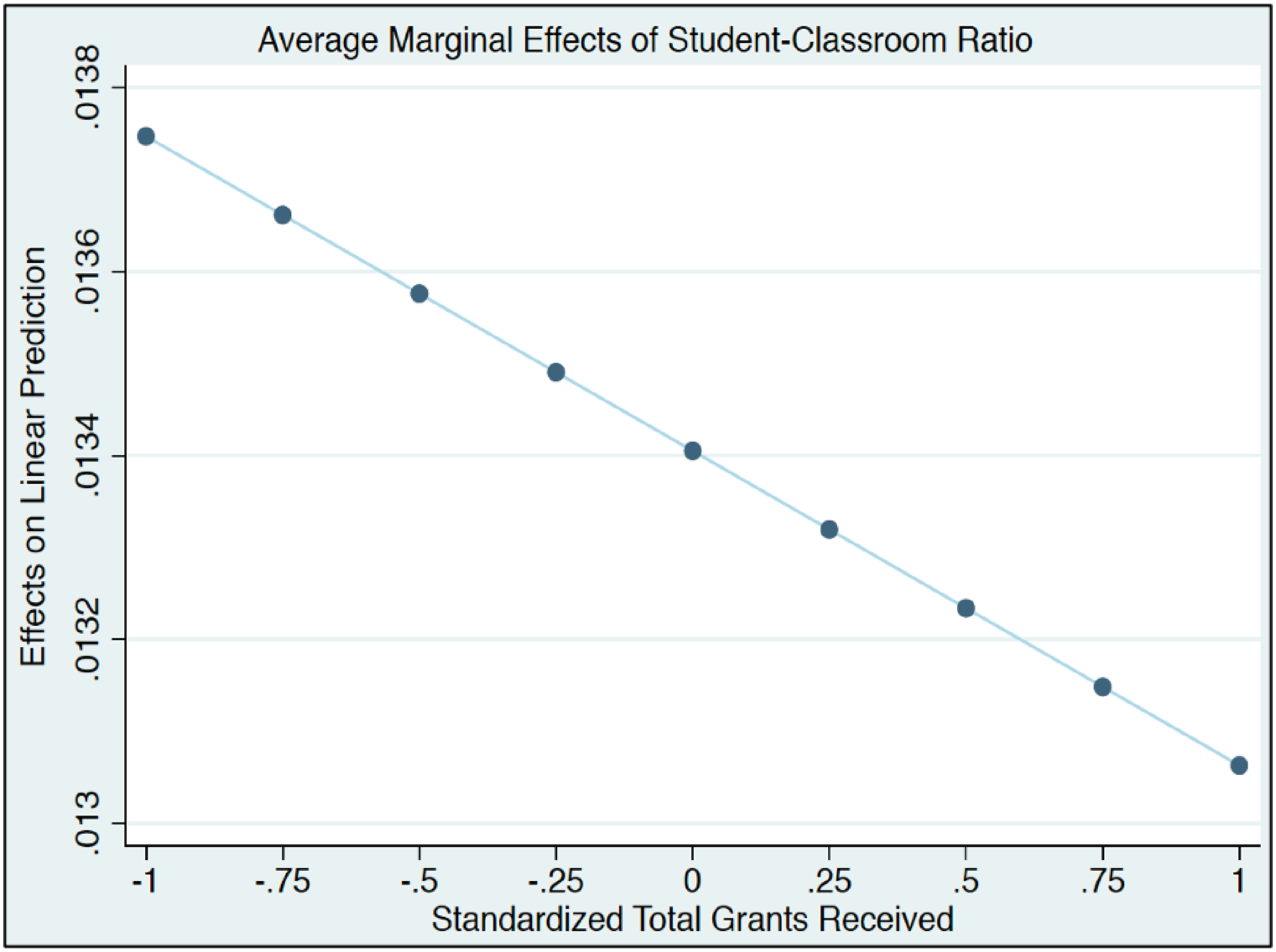
Qualified Teachers x Household Expenditure: Qualified teachers have a positive and constantly increasing marginal effect as household expenditure increases. This is an intuitive result. In districts with higher expenditures (and thus, likely higher incomes), teacher quality is likely also higher (as are school and household resources), explaining the increasing marginal effect. Indeed, the teacher effect might be shadowed by idiosyncratic features and resources of the children studying in wealthier areas. This is suggested by the differences in the trend lines for private and public schools.

Figure 10.6: Marginal Effects: Qualified Teachers x Household Expenditure



Student Classroom Ratio x Grants: We observe positive, but diminishing marginal effects of higher student-classroom ratios on scores as funding increases. The strong marginal effects suggest positive peer effects from larger classroom sizes, which are less pronounced in "wealthier" public schools (but still positive). This might suggest that in schools with low funding, the ability to learn from fellow peers in the absence of school resources might improve performance. However, at well-funded schools, tailored instruction and engagement with school resources might outweigh the positive peer effects.

Figure 10.7: Marginal Effects: Student-Classroom Ratio x Grants Received



10.2.4 Secondary Specification: Fixed Effects & Principal Component Analysis

Table 10.3 (following page) presents the preliminary regression results with fixed effects and principal component analysis, using the total enrollment shares of disadvantaged students enrolled in the school as independent variables.

Table 10.3: Fixed Effects & PCA - Total Enrollment

VARIABLES	(1) Caste Avg. Score	(2) Caste Avg. Score	(3) Religion Avg. Score	(4) Religion Avg. Score	(5) Gender Avg. Score	(6) Gender Avg. Score
Total OBC Enrolled	-0.0371	-0.107***				
	(0.0382)	(0.0372)				
Total SC Enrolled	-0.134***	-0.143***				
	(0.0216)	(0.0206)				
Total ST Enrolled	-0.115***	-0.113***				
	(0.0269)	(0.0252)				
Total Muslim Enrolled			-0.145***	-0.192***		
			(0.0241)	(0.0220)		
Total Girls Enrolled					0.330***	0.300***
					(0.0253)	(0.0262)
Urban vs. Rural	-0.204***	-0.161***	-0.159***	-0.127***	-0.212***	-0.156***
	(0.0414)	(0.0418)	(0.0408)	(0.0408)	(0.0395)	(0.0399)
Private School	1.312***	1.259***	1.389***	1.352***	1.378***	1.303***
	(0.0817)	(0.0803)	(0.0814)	(0.0800)	(0.0803)	(0.0793)
Aided School	0.00473	0.00339	0.00571	0.0187	0.00285	0.00211
	(0.0637)	(0.0649)	(0.0648)	(0.0657)	(0.0633)	(0.0651)
Functional Computers	0.00422***	0.00451***	0.00475***	0.00476***	0.00188	0.00205
	(0.00141)	(0.00137)	(0.00138)	(0.00136)	(0.00142)	(0.00145)
Playground	0.0684	0.0406	0.0646	0.0375	0.102*	0.0755
	(0.0571)	(0.0569)	(0.0569)	(0.0574)	(0.0556)	(0.0567)
Science Lab	0.0692*	0.0891*	0.0662*	0.0817**	0.0462	0.0693*
	(0.0353)	(0.0360)	(0.0354)	(0.0359)	(0.0349)	(0.0359)
Total Grants Received	0.0387	0.0386	0.0354	0.0335	0.0296	0.0307
	(0.0269)	(0.0257)	(0.0259)	(0.0249)	(0.0256)	(0.0250)
Total Contract Teachers	-0.0187	-0.0233	-0.0247	-0.0300*	-0.0161	-0.0176
	(0.0160)	(0.0159)	(0.0167)	(0.0170)	(0.0162)	(0.0163)

Table 10.3 continued from previous page

VARIABLES	(1) Caste Avg. Score	(2) Caste Avg. Score	(3) Religion Avg. Score	(4) Religion Avg. Score	(5) Gender Avg. Score	(6) Gender Avg. Score
Total Qualified Teachers	0.285***	0.316***	0.234***	0.240***	-0.0401	-0.0358
	(0.0384)	(0.0382)	(0.0361)	(0.0354)	(0.0346)	(0.0346)
Total SC Teachers	-0.0589**	-0.0540**	-0.0850***	-0.0795***	-0.0922***	-0.0899***
	(0.0233)	(0.0235)	(0.0234)	(0.0234)	(0.0230)	(0.0233)
Total ST Teachers	-0.0800***	-0.0799***	-0.118***	-0.113***	-0.114***	-0.111***
	(0.0174)	(0.0178)	(0.0121)	(0.0126)	(0.0117)	(0.0121)
Total OBC Teachers	-0.176***	-0.139***	-0.113***	-0.0927***	-0.145***	-0.132***
	(0.0253)	(0.0249)	(0.0237)	(0.0233)	(0.0209)	(0.0211)
Student-Teacher Ratio	-0.000118	-9.67E-05	-0.000172	-0.000148	-0.000140	-0.000181
	(0.000156)	(0.000153)	(0.000155)	(0.000153)	(0.000148)	(0.000149)
Student-Classroom Ratio	0.0189***	0.0203***	0.0174***	0.0176***	0.0101***	0.0113***
	(0.00197)	(0.00197)	(0.00186)	(0.00185)	(0.00184)	(0.00185)
Observations	2,702	2,702	2,702	2,702	2,702	2,702
R-Squared	0.348	0.321	0.342	0.321	0.369	0.334
District FE	YES	NO	YES	NO	YES	NO
PCA	NO	YES	NO	YES	NO	YES

Clustered standard errors in parentheses

*** p <0.01, ** p <0.05, * p <0.1

Columns 1 & 2 construct estimates based on caste; 3 & 4 based on religion; and 5 & 6 based on gender. Columns 1, 3,& 5 incorporate district dummies in the standard fixed effects analysis. Meanwhile, Columns 2, 4, & 6 use principal component analysis to control for other relevant district covariates. By using district-level fixed effects I control for unobserved variation across districts.

Just as in the last regression framework, because the units for most variables are standardized, the general interpretation of any coefficient is as follows: a shift in "x" standard deviations from the mean value of a specific independent variable results in a shift of "y" standard deviations from the mean value of the dependent variable (average test scores). Because there are no interaction terms included in these models, the beta coefficients can be interpreted directly from the regression tables. I conduct these analyses independently of the primary OLS specification, before linking the results in the conclusion.

In (1), a school with an SC enrollment of one standard deviation above the mean (an SC enrollment of 212, rather than the mean of 110), performs .134 standard deviations below the mean of all schools. A school with an ST enrollment of one standard deviation above the mean (an ST enrollment of 75 rather than the mean of 17), performs .115 deviations below the mean. Both estimates are significant at the highest level. In the PCA model (2), the coefficients for SC and ST enrollment remain similar, but the enrollment of OBC children sharply increases and gains significance. A school with an OBC enrollment of one standard deviation above the mean (an OBC enrollment of 1,303 rather than the mean of 643), performs .107 standard deviations below the mean .[8]

Models (3) & (4) suggest a negative relationship between Muslim enrollment and average test score. A school with one standard deviation higher Muslim enrollment has .145 standard deviation lower test scores, compared to the mean. Once more, not all traditionally disadvantaged students are struggling. Schools with higher enrollment shares of girls as compared to boys perform substantially better, according to both (5) & (6). The results from both regressions suggest we may reject our null hypothesis and conclude that schools with higher enrollment shares of disadvantaged castes and religions perform worse than those with lower enrollment shares.

However, this effect appears reversed in the case of gender. Schools with higher enrollment shares of girls perform better than those with lower enrollment shares. It

[8]The size of shifts in standard deviations reflect the substantial amount of variation in the enrollment rates of specific schools. See the descriptive statistics.

is important to note the size of the magnitudes, which are quite small. While there is clearly a robust linear relationship in all three cases, test scores do not appear to change substantially across any of the models, in relation to enrollment shares of disadvantaged groups.

Across all six models, private schools perform substantially better than the excluded category of government schools. These coefficients are of a larger magnitude than the others. However, these estimates do not account for self-selection bias into private schools and cannot control for the inherent talent or resources of wealthier students who are more likely to access these schools. In most models (1-4), the count of functional computers in a school has a positive effect on test score, but with an extremely low magnitude. The presence of a playground or science lab at a school do not seem to relate to higher performance, findings which contrast with the OLS specification. The amount of grant funding (through Sarva Shiksa Abhiyan and Rashtriya Madhavan Shiskha Abhiyan) that a school receives has a slightly positive relationship with test scores, but again, only marginally. In models (1-4), the number of qualified teachers has a positive relationship with student performance. Once more, the counts for teachers from specific disadvantaged castes have negative relationships with student performance. It appears that schools with larger shares of disadvantaged teachers perform worse. This might suggest that these teachers are underprepared, or lack the resources and knowledge required to instruct their students. Lastly, the student to classroom ratio has a positive effect on scores across all models and specifications, suggesting positive peer effects from larger class sizes, or unobserved characteristics of larger schools which allow them to deliver stronger learning outcomes.

Table 10.4 (following page) presents the same estimates, but now uses the percent of students enrolled as the functional form, rather than standardized counts.

Table 10.4: Fixed Effects & PCA - % Enrollment

VARIABLES	(1) Caste Avg. Score	(2) Caste Avg. Score	(3) Religion Avg. Score	(4) Religion Avg. Score	(5) Gender Avg. Score	(6) Gender Avg. Score
OBC % Enrolled	-0.0100***	-0.0102***				
	(0.00130)	(0.00127)				
SC % Enrolled	-0.0184***	-0.0174***				
	(0.00246)	(0.00238)				
ST % Enrolled	-0.0170***	-0.0171***				
	(0.00282)	(0.00287)				
Muslim % Enrolled			0.00645***	0.00765***		
			(0.000823)	(0.000806)		
Girls % Enrolled					0.0130***	0.0128***
					(0.000847)	(0.000859)
Urban vs. Rural	-0.171***	-0.129***	-0.164***	-0.132***	-0.194***	-0.146***
	(0.0410)	(0.0414)	(0.0406)	(0.0405)	(0.0383)	(0.0386)
Private School	1.210***	1.174***	1.497***	1.478***	1.398***	1.330***
	(0.0817)	(0.0805)	(0.0810)	(0.0802)	(0.0798)	(0.0785)
Aided School	-0.0260	-0.0215	0.0159	0.0356	-0.0189	-0.0168
	(0.0623)	(0.0636)	(0.0640)	(0.0650)	(0.0623)	(0.0640)
Functional Computers	0.0039***	0.0034***	0.0035***	0.0029***	0.0035	0.0035
	(0.0013)	(0.0013)	(0.0014)	(0.0013)	(0.0013)	(0.0013)
Playground	0.0412	0.0141	0.0388	0.00793	0.138**	0.112**
	(0.0560)	(0.0571)	(0.0565)	(0.0571)	(0.0552)	(0.0561)
Science Lab	0.0556	0.0725**	0.0608*	0.0747**	0.0556	0.0761**
	(0.0349)	(0.0356)	(0.0351)	(0.0355)	(0.0342)	(0.0351)
Total Grants Received	0.0327	0.0325	0.0342	0.0320	0.0237	0.0242
	(0.0254)	(0.0244)	(0.0261)	(0.0254)	(0.0254)	(0.0246)
Total Contract Teachers	-0.0140	-0.0164	-0.0220	-0.0269	-0.0243	-0.0253
	(0.0166)	(0.0166)	(0.0168)	(0.0170)	(0.0164)	(0.0166)

Table 10.4 continued from previous page

VARIABLES	(1) Caste Avg. Score	(2) Caste Avg. Score	(3) Religion Avg. Score	(4) Religion Avg. Score	(5) Gender Avg. Score	(6) Gender Avg. Score
Total Qualified Teachers	0.965***	0.871***	0.181***	0.170***	0.168***	0.153***
	(0.0302)	(0.0294)	(0.0326)	(0.0316)	(0.0299)	(0.0293)
Total SC Teachers	-0.0577**	-0.0546**	-0.0842***	-0.0765***	-0.0888***	-0.0864***
	(0.0230)	(0.0234)	(0.0233)	(0.0233)	(0.0227)	(0.0231)
Total ST Teachers	-0.0767***	-0.0732***	-0.0975***	-0.0893***	-0.122***	-0.118***
	(0.0164)	(0.0170)	(0.0121)	(0.0128)	(0.0112)	(0.0116)
Total OBC Teachers	-0.0384	-0.0205	-0.107***	-0.0933***	-0.137***	-0.124***
	(0.0273)	(0.0270)	(0.0229)	(0.0225)	(0.0208)	(0.0209)
Student-Teacher Ratio	-0.00023	-0.00021	-0.00022	-0.00020	-4.59E-06	-3.61E-05
	(0.00015)	(0.00015)	(0.00015)	(0.00015)	(0.00014)	(0.00014)
Student-Classroom Ratio	0.0138***	0.0146***	0.0169***	0.0170***	0.0129***	0.0135***
	(0.00188)	(0.00188)	(0.00183)	(0.00182)	(0.00176)	(0.00176)
Observations	2,702	2,702	2,702	2,702	2,702	2,702
R-Squared	0.371	0.343	0.353	0.331	0.386	0.355
District FE	YES	NO	YES	NO	YES	NO
PCA	NO	YES	NO	YES	NO	YES

Clustered standard errors in parentheses

*** p <0.01, ** p <0.05, * p <0.1

Once more, higher shares of disadvantaged groups in a school (by religion or caste) links to lower scores. The results for (2) suggest that for every additional percent of disadvantaged SC enrollment, average test scores are lower by .0102 standard deviations. Thus, a school with an SC enrollment 25% above the mean would have an average score .255 standard deviations lower than the mean. Again, the result is flipped for enrollment by gender. Model (6) suggests that for each additional percent of female enrollment, a given school's scores increase by .0128 standard deviations above the mean.

Tables 10.5 and 10.6 (following pages) use the same functional forms and examine the same relationships as the first two tables, but now restrict the enrollment shares to only Class X students. This serves as a robustness check to the prior estimation: we might expect to see similar magnitudes and signs across the models. Indeed, the results are similar to the estimates in Tables 10.3 and 10.4. The coefficients are slightly smaller in the new regression models, but the estimates remain robust.

The results allow us to make the following conclusions relating to our hypotheses of interest. We may reject null hypotheses (1) and (3), but we fail to reject (2). That is, it appears clear that disadvantaged access links to poor quality in Keralan schools. Schools with higher shares of either minority castes or minority religions (Muslim students) tend to perform worse. However, this does not appear to be the case for schools with larger shares of female enrollment, who tend to perform better. Additionally, private schools perform significantly better than government and aided schools. The result for (2) is mixed, which conforms with the literature. While generally it does not appear that funding or specific input expenditures link to stronger performance, certain inputs make a difference. For example, the number of functional computers at a school is positively correlated with test scores. In certain models, infrastructural developments (playgrounds or science labs) play a small role.

The number of qualified teachers at a school is positively correlated with scores. The Kerala Teacher Eligibility Test (K-TET), distinguishes between four different categories of qualification.[9] Each distinction is based on the qualifications of the exam-taker, and the subject they hope to teach. The data does not allow me to distinguish between qualification types but does suggest that the number of qualified teachers in any category links to stronger scores. Note that most teachers at these schools are qualified, as the qualification is now required by government mandate (under the Right to Education Act).

[9] For the certification process, see here.

Table 10.5: Fixed Effects & PCA - Total Class X Enrollment

VARIABLES	(1) Caste Avg. Score	(2) Caste Avg. Score	(3) Religion Avg. Score	(4) Religion Avg. Score	(5) Gender Avg. Score	(6) Gender Avg. Score
Class X OBC Enrolled	0.0114	-0.0362				
	(0.0293)	(0.0287)				
Class X SC Enrolled	-0.108***	-0.116***				
	(0.0197)	(0.0188)				
Class X ST Enrolled	-0.0945***	-0.0936***				
	(0.0235)	(0.0235)				
Class X Muslim Enrolled			-0.110***	-0.156***		
			(0.0230)	(0.0229)		
Class X Girls Enrolled					0.275***	0.243***
					(0.0233)	(0.0247)
Urban vs. Rural	-0.207***	-0.166***	-0.167***	-0.134***	-0.181***	-0.128***
	(0.0416)	(0.0418)	(0.0408)	(0.0408)	(0.0394)	(0.0399)
Private School	1.348***	1.282***	1.367***	1.313***	1.424***	1.345***
	(0.0816)	(0.0802)	(0.0819)	(0.0801)	(0.0806)	(0.0796)
Aided School	0.0210	0.0209	0.00614	0.0149	0.00242	5.31E-5
	(0.0641)	(0.0653)	(0.0651)	(0.0663)	(0.0632)	(0.0652)
Functional Computers	0.0042***	0.0045***	0.0051***	0.0054***	0.0013	0.0016
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0016)	(0.0016)
Playground	0.0689	0.0424	0.0659	0.0395	0.0917	0.0646
	(0.0570)	(0.0578)	(0.0571)	(0.0577)	(0.0560)	(0.0571)
Science Lab	0.0713**	0.0920**	0.0667*	0.0839**	0.0395	0.0637*
	(0.0355)	(0.0361)	(0.0356)	(0.0361)	(0.0349)	(0.0359)
Total Grants Received	0.0367	0.0375	0.0368	0.0361	0.0281	0.0297
	(0.0266)	(0.0256)	(0.0262)	(0.0252)	(0.0260)	(0.0255)
Total Contract Teachers	-0.0160	-0.0188	-0.0252	-0.0294	-0.0170	-0.0186
	(0.0163)	(0.0162)	(0.0167)	(0.0170)	(0.0164)	(0.0165)

Table 10.5 continued from previous page

VARIABLES	(1) Caste Avg. Score	(2) Caste Avg. Score	(3) Religion Avg. Score	(4) Religion Avg. Score	(5) Gender Avg. Score	(6) Gender Avg. Score
Total Qualified Teachers	0.239***	0.257***	0.223***	0.228***	0.0053	0.0105
	(0.0352)	(0.0351)	(0.0374)	(0.0371)	(0.0328)	(0.0329)
Total SC Teachers	-0.0613***	-0.0570**	-0.0843***	-0.0804***	-0.0932***	-0.0910***
	(0.0235)	(0.0238)	(0.0236)	(0.0236)	(0.0228)	(0.0233)
Total ST Teachers	-0.0768***	-0.0762***	-0.120***	-0.116***	-0.116***	-0.113***
	(0.0199)	(0.0207)	(0.0124)	(0.0130)	(0.0114)	(0.0119)
Total OBC Teachers	-0.183***	-0.157***	-0.125***	-0.105***	-0.155***	-0.140***
	(0.0236)	(0.0232)	(0.0234)	(0.0233)	(0.0206)	(0.0209)
Student-Teacher Ratio	-0.00014	-0.00013	-0.00017	-0.00014	-0.00018	-0.00022
	(0.00016)	(0.00016)	(0.00016)	(0.00016)	(0.00014)	(0.00014)
Student-Classroom Ratio	0.0178***	0.0185***	0.0169***	0.0170***	0.0138***	0.0147***
	(0.00188)	(0.00189)	(0.00187)	(0.00187)	(0.00181)	(0.00182)
Observations	2,702	2,702	2,702	2,702	2,702	2,702
R-Squared	0.345	0.318	0.340	0.316	0.364	0.328
District FE	YES	NO	YES	NO	YES	NO
PCA	NO	YES	NO	YES	NO	YES

Clustered standard errors in parentheses

*** p <0.01, ** p <0.05, * p <0.1

Table 10.6: Fixed Effects & PCA - % Class X Enrollment

VARIABLES	(1) Caste Avg. Score	(2) Caste Avg. Score	(3) Religion Avg. Score	(4) Religion Avg. Score	(5) Gender Avg. Score	(6) Gender Avg. Score
Class X OBC % Enrolled	0.00882***	0.00905***				
	(0.00106)	(0.00104)				
Class X SC % Enrolled	-0.0158***	-0.0148***				
	(0.00192)	(0.00186)				
Class X ST % Enrolled	-0.0150***	-0.0153***				
	(0.00266)	(0.00272)				
Class X Muslim % Enrolled			0.00638***	0.00740***		
			(0.00206)	(0.00075)		
Class X Girls % Enrolled					0.0110***	0.0108***
					(0.000717)	(0.000732)
Urban vs. Rural	-0.171***	-0.127***	-0.165***	-0.134***	-0.171***	-0.121***
	(0.0409)	(0.0414)	(0.0758)	(0.0404)	(0.0380)	(0.0383)
Private School	1.208***	1.171***	1.485***	1.460***	1.376***	1.306***
	(0.0814)	(0.0804)	(0.154)	(0.0795)	(0.0802)	(0.0788)
Aided School	-0.0101	-0.00798	0.0178	0.0371	-0.0216	-0.0194
	(0.0623)	(0.0638)	(0.121)	(0.0651)	(0.0624)	(0.0641)
Functional Computers	0.0036***	0.0032**	0.0035**	0.0030**	0.0031**	0.0032**
	(0.0014)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
Playground	0.0474	0.0194	0.0406	0.0110	0.133**	0.108**
	(0.0561)	(0.0571)	(0.0719)	(0.0571)	(0.0550)	(0.0558)
Science Lab	0.0547	0.0722**	0.0597	0.0736**	0.0544	0.0746**
	(0.0347)	(0.0354)	(0.0414)	(0.0355)	(0.0343)	(0.0351)
Total Grants Received	0.0351	0.0352	0.0345	0.0325	0.0231	0.0237
	(0.0259)	(0.0250)	(0.0258)	(0.0256)	(0.0257)	(0.0248)
Total Contract Teachers	-0.0159	-0.0174	-0.0210	-0.0258	-0.0225	-0.0235
	(0.0163)	(0.0165)	(0.0196)	(0.0170)	(0.0168)	(0.0169)

Table 10.6 continued from previous page

VARIABLES	(1) Caste Avg. Score	(2) Caste Avg. Score	(3) Religion Avg. Score	(4) Religion Avg. Score	(5) Gender Avg. Score	(6) Gender Avg. Score
Total Qualified Teachers	0.0943***	0.0853***	0.180***	0.169***	0.170***	0.154***
	(0.0296)	(0.0288)	(0.0514)	(0.0316)	(0.0301)	(0.0294)
Total SC Teachers	-0.0609***	-0.0587**	-0.0855**	-0.0781***	-0.0853***	-0.0836***
	(0.0230)	(0.0234)	(0.0381)	(0.0233)	(0.0225)	(0.0228)
Total ST Teachers	-0.0827***	-0.0785***	-0.0977***	-0.0903***	-0.123***	-0.119***
	(0.0162)	(0.0169)	(0.0122)	(0.0126)	(0.0117)	(0.0121)
Total OBC Teachers	-0.0552**	-0.0364	-0.106**	-0.0927***	-0.141***	-0.127***
	(0.0249)	(0.0247)	(0.0398)	(0.0223)	(0.0209)	(0.0210)
Student-Teacher Ratio	-0.00022	-0.00021	-0.00020	-0.00018	-1.49E-06	-4.75E-05
	(0.00015)	(0.00015)	(0.00020)	(0.00015)	(0.00014)	(0.00014)
Student-Classroom Ratio	0.0147***	0.0155***	0.0170***	0.0172***	0.0137***	0.0143***
	(0.00184)	(0.00185)	(0.00220)	(0.00182)	(0.00179)	(0.00179)
Observations	2,691	2,691	2,691	2,691	2,691	2,691
R-Squared	0.374	0.345	0.356	0.334	0.388	0.356
District FE	YES	NO	YES	NO	YES	NO
PCA	NO	YES	NO	YES	NO	YES

Clustered standard errors in parentheses

*** p <0.01, ** p <0.05, * p <0.1

Additionally, schools with higher numbers of (qualified) disadvantaged teachers perform worse. This might suggest that structural inequalities in learning outcomes extends across an individual's lifecycle. That is, disadvantaged castes who struggle in primary and secondary education but earn a qualification from the government to teach might be less able to instruct their students, even if they hold a shared social identity. Of course, all of these relationships might be linked to confounding factors in the external environment, such as district-level features of income, literacy, or cost of living. My regressions, which control for many of these socioeconomic features in the PCA analysis, and control for a school's location within a district in the standard fixed effects, alleviate some of these concerns. A series of checks using variance inflation factors determined that the models do not suffer from high collinearity, and the clustered standard errors account for potential heteroskedasticity.

10.2.5 Limitations

Before turning to policy recommendations and motivating future research, it is important to consider some of my limitations, which mirror those of other non-experimental studies in the literature. The first of these is a lack of data below the subdistrict level. This is particularly relevant for socioeconomic controls, which were primarily scraped at the district level due to sparse collection at the panchayat or block level. However, I am able to contribute to the existing nonexperimental literature by gathering data on learning outcomes at the level of the schools themselves, a more granular level of analysis. Other data sources on learning outcomes, including ASER, the IHDS, and the NASS either do not provide interpretable estimates below the district level, or do not have substantially large sample sizes.

As with the school report cards, most district covariates were not collected in the same year as the SSLC examination (2019). This could create bias if the included controls have drastically changed over the period from their collection to the test date. However, the earliest data was collected from 2015, and most data was collected from 2017-2018. Regardless, there is potential for omitted variable bias across the specifications. In particular, I was unable to gather data on the students themselves, their families, or the neighborhoods in which they live.

The descriptive analysis suggests the lack of sizeable variation across certain indicators (including the scores themselves). Because of the low sample variance, the estimators and p-statistics in the regression may be miscalculated. This, along with UDISE's own measurement error and sample selectivity, suggest the results must be interpreted cautiously.

As mentioned before, the large private school effect is likely masked by self-selection of children into these schools. The results are almost certainly over-estimates of true student performance within these settings, especially if these schools cater to the elite (the demographic composition of the schools suggest this is not such an obvious conclusion to make). There is also selectivity into the sample itself. The literature and Keralan government suggest a large portion of Scheduled Tribe and Scheduled Caste students drop out of school prior to the SSLC examinations (TP, 2019). The data reflects these findings. In Wayanad, the district with the highest population share of Scheduled Tribes (see Figure 2); the sample has the smallest count of schools taking the examination (only 38 schools; or 1.4% of the sample).

Another limitation relates to the use of test scores themselves. Some scholars argue that state standardized testing is not truly reflective of students' cognitive skills development, because of the filtering system of the curricula and focus on rote memorization. However, Kerala, a leader in educational outcomes, has a renowned system and a well-defined curriculum. Furthermore, even if independent assessments of researchers or ASER surveys provide better estimates of skills development, they are either not standardized or are unavailable at the school level. This thesis focuses on the performance of disadvantaged groups within this system of filtration, to examine whether or not they truly lag behind. Whether or not state sponsored education is crafted to develop learning outcomes, the results do suggest that the learning experience is different for disadvantaged groups, even in the "best case".

Because this study is restricted to the state of Kerala, it is not externally valid to the whole of India. However, the intent of the thesis is not to provide a wholistic picture of Indian education, but rather test the theories of systemic inequality using a case study from the "best" possible case of learning outcomes in the nation. Finally, as this is a nonexperimental study, I cannot claim any degree of causality. Regardless, these observed relationships may guide future studies or RCTs which are able to explicitly create a causal mechanism linking learning outcomes to my covariates of interest.

Chapter 11

Conclusions

The regression results present a mixed picture. On the one hand, I may reject my null hypotheses for conjectures **(1)** and **(3)**. Higher SC, ST, and Muslim enrollments all link to lower test scores, after accounting for controls. Not only do most disadvantaged groups (with the exception of females) perform worse, but private schools perform substantially better (even those with high disadvantaged enrollments). Even without government aid, and despite Kerala's strong state government, private schools perform better than public institutions in Kerala. Aided schools also perform better than government schools, despite receiving fewer grants. Thus, the results support the overarching theory of education as a tool for elite domination, rather than a platform for liberation. Such a result in the "best case" of Indian state education is troubling.

On the other hand, in the case of hypothesis **(1)**, the magnitudes of the coefficients are relatively small. While the results are statistically significant, they do not appear practically significant at scale. Because there is such a large degree of variation in the distribution of student enrollments, a one deviation increase relates to hundreds of additional student enrollments. For example, an increase of 100 SC students relates to a loss of only .134 standard deviations in a school's test scores. Additionally, most schools have comparable scores (at the mean) and perform exceptionally well on state examinations. In 2018, 97% of students who sat Kerala's SSLC passed the examination. Hence, even disadvantaged students who are "pushed" into Keralan education are well prepared for examinations. But this may mask important deviations in the results of those who passed. Given the educational excellence of Kerala, such a result is more meaningful than it might appear. If most students perform well on these examinations (i.e. pass), and these examinations are indeed used as filtration systems to "rank" students relative to one another, then even one tenth of a standard deviation can be the difference between a high school education and admittance to a top university.

When we examine school type based on access, rather than performance, the results align with the literature on education as a platform for liberation. We observe this in the high enrollments of OBC and Muslim students in private schools (particularly girls). This suggests that parents across these minorities are willing to invest in their daughters' education unlike elsewhere in India. Additionally, girls-only schools and those with higher female enrollments tend to outperform coed and boys-only schools. Once these girls are included in Keralan education, they succeed. However, these results are far less optimistic for SC and ST minorities. The literature suggests many students of these castes drop out before the 10th standard. We directly observe this in the enrollment data, and I note that this selectivity influences the results of the analysis. For students from these backgrounds who do continue to attend school, there are higher enrollments for these groups in coed and boys-only schools, suggesting that in these groups, parents either make the "education investment" for boys in the family over girls, or that girls have external commitments to family (perhaps as a caretaker) which influence the decision to drop out.

The mixed results from hypothesis **(3)** largely conform with the literature on input-based education but motivate further research to determine exactly which bundles of inputs foster learning outcomes, as Mbiti and his co-authors recently explored. For example, while the presence of certain inputs (such as the number of working computers at a school or whether they have a playground) seem to link to stronger learning outcomes, others (such as the grant funding received or whether or not the school has a science lab) are less important. I suggest the state government shift from providing cash transfers to purchasing specific input bundles for schools in need. However, as the "One Laptop per Child" initiative demonstrated, input purchases alone will not result in stronger outcomes (Robertson, 2018). These big push initiatives must be bundled with extensive spending on teacher recruitment and training. Household expenditure on schooling (in the absence of school funding) is also positively correlated with learning outcomes; therefore, government campaigns promoting educational investment (particularly for ST and SC groups) might be beneficial. Targeting gender norms in these marginalized communities is also likely to produce beneficial results for girls currently excluded from public education.

Examining pedagogy rather than inputs provides a clearer picture of Kerala's future direction. The number of qualified teachers at a school has a sizable positive relationship with test scores. This suggests that placing funding in teacher recruitment and training is a cost-effective strategy to improve performance. The student-classroom ratio also holds a positive relationship with test scores, suggesting positive peer effects in larger classrooms. This is not the case across all of India. There are many studies that document how large classrooms exacerbate learning gaps for the students who do not receive instruction. However, in combination with technology-aided instruction, learning from peers might complement individualized instruction (Mitra, 2012). Rather than focusing the majority of funding on inputs and infrastructure, Kerala would be well-advised to prepare candidates for the teacher qualifying exam (K-TET) and train current teachers on how to correctly implement technology-aided instruction.

The data also suggests that there are barely any SC and ST teachers hired across school types. This reflects educational inequalities throughout the lifecycle of these individuals. To begin, these individuals are far more likely to drop out prior to the 10th standard (date of the SSLC examination), which we observe in the enrollment data. Because of this, they are unable to receive a secondary school leaving certificate or a basic college education. The K-TET, which requires these certifications as prerequisites to become a qualified teacher, automatically eliminates a large number of teachers from disadvantaged castes from pursuing a teaching career.

These indirect exclusions mandated by K-TET seem validated by the findings. The regression analysis suggests there is a negative relationship between the count of disadvantaged teachers hired by a school and that school's test scores. However, when accounting for marginal effects in the interaction terms, we observe that at schools which consist of a large disadvantaged student base, the marginal effects of disadvantaged teachers from the same caste, while negative, are increasing. This suggests the training and recruitment of disadvantaged teachers is particularly important. I suggest shifting well-performing students from poorer backgrounds into the teaching track from a young age, before the SSLC examination itself. By learning the curriculum and receiving teacher support, these students may become better able and motivated to pursue teaching in the future.

In terms of school choice, as mentioned earlier in the thesis, the correct strategy is not to prioritize one sector of education over another, but instead attempt to integrate the two into a collaborative system. This is an incredibly difficult task, both because of parents who place their children in higher-performing private schools and thus shift their funding, and because of teachers and coalitions at public schools, who continue to demand more government funding. Kerala should identify and target the lowest achievers, to create a level playing field. It is clear from the regressions that, self-selection or not, private schools perform far better than public learning institutions. This despite various state programs bolstering public education, and the Union's own federal funding through RMSA and RTE. However, the state is focusing primarily on improving access, rather than quality in these public institutions. Infrastructural investments are more visible and thus politically palatable to constituents, but teacher training programs and technology-aided instruction are far more likely to improve outcomes.

In order to level the playing field, I present a policy recommendation suggested by Karthik Muralidharan (Muralidharan "Priorities for Primary Education, 2013). Muralidharan advocates for afterschool programs which are targeted towards students who have fallen behind the curricula. Private tutoring paid for by wealthy parents for their children is a large source of school expenditure in Keralan districts. However, large-scale tutoring programs funded by the government to accompany public instruction would help those without access to private tutors. These programs need not be serviced by qualified teachers, but rather by individuals familiar with the school curricula and the communities in which they work. This would alleviate learning gaps in the student base and could reduce dropout rates for disadvantaged students. If this tutoring base, which caters towards disadvantaged minorities, also consists of community members from these castes, the marginal effects analysis suggests these group-based seminars could produce positive peer and "mentor" effects for ST and SC students.

The geographic distribution of scores and funding across districts appears relatively uniform, which suggests the state government is allocating resources equitably across districts and school boards. However, certain rural clusters which house a larger share of disadvantaged minorities (such as the district of Wayanad), have noticeably fewer schools and lower performance than other districts. The state government needs to delegate more funding to the programs suggested above by targeting clusters of ST and SC groups, in order to integrate them within the schooling system and ensure lower dropout rates in elementary and lower secondary education.

The results of my thesis and these policy recommendations generate a need for further research. The largest limitation of this study is a lack of data availability in the Indian context. However, with a longer timescale and a more directed collection procedure, researchers should be able to generate a panel dataset of my covariates of interest. An ideal panel study would extend the analyses of scholars such as Professors Raj Chetty and Nathaniel Hendren, who map the (in)equality of opportunity in higher education.[1] Professors Chetty and Hendren extend my analysis of intergenerational mobility to later life outcomes. By tracking individuals across their lifecycle, Professor Chetty is able to model how an individual's educational attainment and performance influences their later career trajectory. Using this data, he can map differences in outcomes between disadvantaged minorities and elite groups. His research currently indicates that, in the American context, the outcomes of dis-

[1] See Opportunity Insights for the research strands of Professors Chetty and Hendren.

advantaged minorities (i.e. first-generation students, low-income students, women, or racial minorities) do not substantially differ once they gain access to top institutions. However, the degree to which these institutions offer access to these groups greatly varies.

Professors Chetty, Friedman, Saez, Turner and Yagan construct indices of mobility rates, or the product of the percentage of students at colleges from the lowest quintile of the income distribution and the percentage of students from this same quintile who go on to place within the top quintile of the income distribution later in life (literally, a "rags to riches" index). The authors find incredible heterogeneity in the mobility rates of college campuses across the United States, meaning that not only do access rates vary, but so too do later life outcomes conditional on access. Ivy League institutions fall near the bottom of these mobility rate constructs, which is perhaps unsurprising given the history of their establishment.[2] Meanwhile, "mid-tier" public institutions have higher mobility rates, though fewer students place in the top 1% of the income distribution (Chetty et. al 2017).

This heterogeneity emphasizes the need for a structural policy or explicit requirement of inclusion. Professor Chetty's research strand links also analyzes heterogenous impacts across population subgroups; but establishes causal links. I do not yet have the technical ability nor the endowments to pursue this ambitious research, but this project is the first step in this direction. Below, I review the most relevant covariates for future research which were excluded from this study, at the level of the neighborhood, the family, and the child.

At the level of the neighborhood, spatial analysis could isolate community-level (panchayat) factors linking to educational attainment and performance. For example, proximity to government offices might indicate a higher level of income and stronger investment by public officials for proximate constituents. The presence of universities could mold individual cost-benefit analyses relating to the value of education.

[2]For more context, see the following interview of Dr. William Deresiewicz, the author of "Excellent Sheep".

In areas with higher returns to human capital and knowledge accumulation, communities are more likely to invest in education. At the family level, the single most relevant covariate of interest is income. Professor Chetty's studies on Ivy League institutions in the United States are undoubtedly externally valid to other global contexts. Indeed, in a filtration system such as India, familial investments in private tutoring, connections with community leaders, or endowments which allow for access to additional inputs are even more likely to influence outcomes.

Furthermore, while I have managed to control for some elements of social identity, the data did not allow for me to control for many other relevant indicators of familial support or stability. I do not, for example, have information on the educational attainment of the parents, their investments in their children's education, or their occupations. Lastly, at the level of the child, I cannot control for inherent differences in idiosyncratic talent or ability, prosocial motivation, or psychosocial indicators. All of these factors are likely to establish a stronger (causal) link between enrollment and learning outcomes.

Overall, the conclusions from this thesis suggest that even in the context of Kerala, the "best case" of Indian education, there remains ample room for improvement. The exceptional pass rates of disadvantaged students despite their poor relative performance, the inclusion of certain disadvantaged minorities (OBC and Muslim) in private education, and the equitable allocation of resources across most districts, suggests Kerala is doing something right. At the same time, there are significant negative relationships between a school's learning outcomes and the shares of their disadvantaged students and teachers. This relationship remains robust across functional forms and estimation methods. Policies should be crafted which improve cognitive understanding for both disadvantaged students and provide training for disadvantaged teachers to better instruct their students on the curriculum.

Although the state may be misplacing funding into specific inputs or infrastructural developments, Kerala's effort to integrate students into their state system in line with Sarva Shiksa Abhiyan and Rashtriya Madhyamik Shiksa Abhiyan appears generally successful. Large portions of disadvantaged students are enrolled in secondary education, and although they may not perform favorably compared to their elite peers, they continue to pass their examinations. This might be due to a number of confounding factors, including features of these disadvantaged children themselves (i.e. idiosyncratic talent or endowments), their families, or their neighborhoods.[3] More likely, institutional features such as the strength of state education in Kerala and the state's historical emphasis on equality and inclusion play a stronger role. Determining the true effect of Kerala's institutional features on learning outcomes for the disadvantaged is beyond the scope of this thesis but should motivate further research.

This thesis has contributed to scholarship in education economics by testing a wide range of sub-hypotheses across the literature and integrating these metrics into a single study. It is one of the only nonexperimental papers to present estimates at the level of the school, the smallest possible unit of analysis outside of a Randomized Control Trial and tailored surveys. By merging UDISE report card data with the associated test scores of the schools, this is the first study to create a comprehensive dataset linking government-sponsored state examinations to features of the universe of secondary schools. This is also one of the only papers examining the context of Kerala, motivated by the consensus that Kerala leads the pack in educational outcomes. The correlations of interest strongly motivate future study of disadvantaged performance across all contexts of India, not just the poorly performing states.

[3] Both for those enrolled, and for those who dropout prior to Class X and are thus unobserved in the sample.