Caste Discrimination in School Admissions: Evidence from Karnataka, India

Alaka Holla*
Brown University

JOB MARKET PAPER

22 January 2007

Abstract

Scheduled Castes and Scheduled Tribes (SC/ST), or groups historically labeled "untouchables" and tribals, tend to cluster in the worst secondary schools in India. This paper tests whether this clustering results from caste-based prejudice in admissions or from schools' objective to maximize test-scores on a state-wide matriculation exam. It uses test-score data from the state of Karnataka, where no administrative boundaries limit an individual's choice of schools, around 65% of schools fall under private management, and no affirmative action policies guarantee minimal levels of representation in secondary school. If schools were discriminating against SC/ST applicants, they would set higher admissions standards for them in terms of test-score potential demonstrated at the time of admission. On the matriculation exam administered three years later, SC/ST students would then outperform their non-SC/ST counterparts within the same school. The results from a sample of roughly 5.7 million test-takers from the full universe of secondary schools in the 1996-2004 period are consistent with the hypothesis that schools do not exhibit prejudice against SC/ST students in admissions. As a group, SC/STs do not outperform their non-SC/ST classmates, and this holds both when differences in mean performance proxy for differences in admissions thresholds and when various points (quantiles) of the SC/ST and non-SC/ST test-score distributions are compared. This result also holds along the school performance distribution within attendance areas and when the analysis is limited to private schools.

*I thank Kaivan Munshi, Andrew Foster, and David Weil for all of their guidance and encouragement. I am also grateful to Eswar Bhat and Chamaraja Andini of the Karnataka Secondary Education Examination Board for providing me with data, Spatial Data Private Limited for mapping Karnataka's secondary schools, and the staff of Karnataka's Department of Public Instruction and P.S Karanth for answering numerous queries. This paper also benefited from comments from Kevin Lang, Glenn Loury, Malhar Nabar, Yona Rubinstein, Ivo Welch, and participants of Brown's Micro Lunch and Applied Microeconomics Seminar. I am responsible for any errors that remain.

1 Introduction

Discrimination in access to educational resources continues to be a policy concern in industrialized countries that have experienced substantial civil rights movements, but there have been few quantitative studies in developing countries, where stratification along ethnic and caste lines is just as severe, if not more. The caste system in India, for example, has long dictated the scope of social interactions among groups, restricting marriages, the use of public spaces and temples, and even the sharing of food and water. Previously referred to as "untouchables" and tribals, those groups identified by the Indian Constitution in 1950 as Scheduled Caste or Scheduled Tribe (SC/ST) sit at the bottom of the caste-hierarchy. To redress centuries of social disadvantage, there are a number of affirmative action policies that mandate minimum levels of SC/ST representation in local and state governments, universities, and public-sector employment. These reservations or quotas, however, generally do not apply to secondary schools.[1] Given the degree of interactions that take place within a school and given the absence of policies limiting segregation by caste, would SC/ST students be able to access the same high schools as non-SC/ST students? Would they face discrimination in high school admissions?

This paper develops and implements a test for caste discrimination in secondary school admissions, using administrative test-score data from 1996-2004 from a state-wide matriculation exam administered in the Indian state of Karnataka. In this state, no administrative boundaries or residence requirements limit an individual's choice of schools, and over 65% of high schools are privately managed.[2] Given this institutional setup and the absence of affirmative action policies, if there were taste-based discrimination against SC/STs in the sense of Becker (1957) - that is, if non-SC/STs were averse to interacting with them in the context of a school - then schools and students should be able to act on these preferences, and there should be evidence of this in the data.

In addition to Karnataka's institutional setup in education, the data used in this paper - namely, the scores of the full universe of test-takers from the Karnataka Secondary Education Examination Board (KSEEB) and GIS data on the location of schools - allows the proposed test for discrimination to circumvent many issues of measurement that have plagued the discrimination literature. First, the test-scores themselves allow the analysis to go beyond a simple "bench-marking test" or segregation

[1] One exception to this rule can be found in *Kendriya Vidhyalayas* - centrally administered schools that cater to the children of government employees.

[2] High School Database, Department of Public Instruction, Government of Karnataka, 2002.

index and avoid empirical specifications vulnerable to omitted variables bias. In a bench-mark test or segregation index, the shares of SC/STs within schools would be bench-marked against their shares in the population or would be compared across schools. This, however, would not address the issue of sorting along other relevant dimensions - in the case of schools, sorting by academic ability. Figure 1, for example, demonstrates that, as in other states of India, SC/STs in Karnataka are generally consigned to the worst schools. It ranks schools by their average non-SC/ST performance on the 2004 matriculation exam and shows the average SC/ST composition of high schools in Karnataka in each quintile of the school-performance distribution alongside the average SC/ST share of the villages in which these schools operate.[3] SC/ST students cluster in the schools in the bottom quintiles - a concentration that cannot be fully explained by the placement of poor quality schools in predominantly SC/ST areas.

SC/STs, however, also tend to come from poorer socio-economic backgrounds than other castes. Table 1 displays statistics from the 2001 Census of India on education completion rates and asset ownership that indicate a considerable degree of inter-caste inequality among Karnataka's adult population. It is possible that this inequality in the previous generation and the consequent inequality in human capital investment prior to high school in the current generation account for the observed allocation of SC/ST students to the worst schools. The good schools could simply be trying to maximize their average performance on the exam and therefore denying admission to SC/ST students who enter high school less prepared

An empirical strategy that regresses the probability of going to a certain school on caste and a vector of observable individual characteristics such as parental income or education could also deliver a spurious discrimination result since it presumes that there are no other variables, unobservable to the researcher but observable to whoever makes school admissions decisions, that could be negatively correlated with caste. Such a criticism would be similar to one which Becker (1993) levied against the Boston Fed Study on racial discrimination in mortgage lending by Munnell et al. (1992), which found that in a regression with the probability of receiving a loan as the dependent variable, an indicator for being an African American remained significantly negative even after controlling for an array of individual characteristics that could proxy for credit risk. Becker (1993) argued that a cleaner test, one that would suffer less from omitted variables bias, would involve examining the

[3]If SC/ST students tend to perform worse than non-SC/ST students, then schools at the bottom quintiles might be performing poorly simply because they have high SC/ST compositions. To ensure that a skewed allocation does not represent a tautology, Figure 1 uses only the performance of non-SC/ST students to divide schools into quintiles.

degree of selection exhibited by the group that potentially faces discrimination - specifically ascertaining whether African Americans who did receive loans displayed lower default rates relative to successful white applicants, which would suggest that they had been held to stricter standards. This is also the reasoning behind Karabel's (2005) claim that the Ivy League schools discriminated against Jewish applicants when he cites an internal report at Harvard in the 1950s that showed that their Jewish students, who were subject to implicit admissions quotas since the 1920s, outperformed their gentile peers in terms of both their studies and their behavior; and behind conclusion that there were entry barriers for black players in major league baseball in 1967 by Pascal and Rapping (1972) who show that black players of various positions demonstrated higher lifetime batting averages than their white counterparts.

This is also the simple intuition behind the test of caste discrimination in this paper, which examines the degree of selection exhibited by SC/STs *within schools*. If a school only cares about maximizing its test scores and shows no preference for a certain caste-composition and if it can perfectly screen students by their test-score potential, then within the school admissions thresholds in terms of this test-score potential should be the same across castes. As a result, the marginal SC/ST and non-SC/ST students should perform identically on the exam administered three years after admission. If, however, a school also exhibits a taste for discrimination in the sense of Becker (1957) and tries to limit its SC/ST share, then admissions criteria should be stricter for SC/ST students and the marginal SC/ST student should outperform her non-SC/ST counterpart within the school. When schools cannot perfectly screen students by their test-score potential and must use caste as a predictor of test performance, allocations in which the marginal SC/ST and non-SC/ST students perform identically could result from statistical discrimination (Arrow, 1973). A simple model will make clear that allocations in which the marginal SC/ST student does not perform better than her non-SC/ST counterpart are consistent with taste-based discrimination against SC/STs in admissions only if schools are also simultaneously *overestimating* the test-score potential of SC/STs.

To test for differential admission cutoffs across castes in terms of test-score potential, this paper uses individual-level test-score data from the full sample of high school students who took the matriculation exam from 1996-2004 (roughly 5.7 million test-takers). This data permits a comparison of SC/ST and non-SC/ST distributions within schools - not only with respect to their means but also with respect to other points of the distributions, making the test of differential thresholds to some extent independent of assumptions on the distributions of test-score potential within each caste-group.

In their paper on racial bias in motor vehicle searches, Anwar and Fang (2006) discuss how papers in the discrimination literature must often make a number of assumptions on behavior in order to conflate the *marginal* (the focus of theory) and the *average* (what can be observed in the data) when the underlying distributions of interest are unobserved (credit riskiness in the case of mortgage lending, propensity to carry drugs in the case of motor vehicle searches) and when the the observed outcomes are dichotomous (loan default, presence of drugs). The individual-level data used in this paper along with the assumption that the test-score production process is monotonic with respect to a students' unobserved incoming test-score potential allows a more accurate measurement of the difference between the marginal SC/ST and non-SC/ST students within a school.

Further, information on the location of schools and a geo-coded map of Karnataka's villages helps to define distance-based school-attendance areas, pinpoint the complete set of schools that an individual could in principle access, and rank the schools within this set. Testing for discrimination along the entire distribution of school performance counters concerns that discrimination against SC/STs in one part of the distribution of school quality could be obscured by favorable treatment in admissions in another when averaging across all schools and provides an additional way of addressing the possibility that SC/STs and non-SC/STs participating in secondary school are drawn from different distributions of test-score potential.

The results of this empirical exercise suggest that caste discrimination in school admissions *cannot* explain why SC/ST individuals are consigned to poor performing schools. The caste gap in terms of performance narrows considerably when moving from comparisons of the two groups in the population to within-school comparisons, suggesting that ability acquired prior to admission steers the allocation of students to schools to a large extent. The gap, however, never reverses itself. SC/ST students do not face stricter admissions thresholds than their non-SC/ST classmates. This result holds in a flexible quantile framework for all points of the SC/ST and non-SC/ST distributions within schools; in the best, worst, and intermediate schools in school-attendance areas; in urban areas where there is an abundance of schools to choose from and in rural areas where school-choice is more limited; in areas with low shares of SC/ST individuals and in areas where they are less of a minority.

This absence of taste-based discrimination against SC/STs in the sphere of education, however, departs from evidence in the US literature that suggests individuals prefer to go to school with members of their own race. Boustan (2006) finds that even after accounting for the correlation between race and income in 1970, homeowners were willing to pay a premium to live in areas in which the student population of the nearest local public high school was predominately white. Similarly, in investigating the size of current political jurisdictions, Alesina, Baqir, and Hoxby (2004) show that the number of school-attendance areas in a county responds to increases in racial heterogeneity even after controlling for the number of districts and the degree of income heterogeneity, a result which the authors take as suggestive evidence of individuals' wanting to avoid interactions with members of another race in school.

These test-score patterns in Karnataka's high schools have important implications for a broader, ongoing debate in India about the merits of extending affirmative action policies to primary and secondary schools. Policies to address inter-caste inequality continue to be a divisive issue in India as recent protests against reservation hikes in universities have demonstrated. Anti-affirmative action arguments often propose that the the government change the way it targets beneficiaries and move to more individual-based indicators such as parental income or education. Those in favor of reservations counter that group-based indicators matter since caste discrimination still limits the ability of certain groups to take advantage of opportunities for social mobility, regardless of their income. To the extent that similar patterns would be found in primary schools, the findings of this paper give statistical support to the view that interventions aiming to increase SC/ST access to school quality during secondary school *need* not target individuals on the basis of their caste.

The next section provides some background on caste and the institutional setting of secondary education in Karnataka. Section 3 presents a simple model of sorting that delivers testable predictions on test-score patterns that distinguish allocations arising from sorting based solely on test-score potential from those that involve an additional caste component. Section 4 contains the empirical analysis: it describes the data used to test these predictions and presents the results of the empirical specifications suggested by the model. Section 5 concludes.

2 Background and institutional setting

2.1 Caste and education

The potential for caste-discrimination specifically in the sphere of education dates back to ancient texts that prescribed proper codes of conduct for members of each caste, some of which still inform religious rituals practised today. The *Vedas* - a compendium of scientific knowledge, ethics, religious hymns, and prayers - served as the primary texts of instruction, and a *Sudra* was a member of a "low-caste" community and a step *above* an "untouchable". The *Gautama Dharama Sutra* warns,[4]

> Now if a Sudra listens intentionally to (a recitation of) the Veda, his ears shall be filled with (molten) tin or lac.
> If a Sudra recites (Vedic) texts, his tongue shall be cut out.
> If he remembers them, his body shall be split in twain.

Among the situations that should trigger a suspension of Vedic recitation, the *Manu Smriti* lists proximity to a corpse, the sound of weeping, and the presence of a *Sudra*.[5]

During the colonial period, Christian missionaries and the British colonial authorities started separate schools for "untouchables" and tribal people, after realizing that other communities had barred them from participating in the public school system (Srinivas (1968), Jaffrelot (2006)). After Independence, Articles 15 and 17 of the Indian Constitution prohibited discrimination on the grounds of caste, and an amendment in 1951 specifically sanctioned states' use of reservations in schools and colleges on the grounds of caste. While Karnataka's reservations in higher education guarantee SC/ST students 23% of slots in public universities,[6] SC/ST-specific policies applicable to primary and secondary education have mainly consisted of subsidies for fees, uniforms, and books and the provision of mid-day meals (Department of Public Instruction, Karnataka, 2003).

Although it would be difficult to distinguish SC/ST and non-SC/ST students by their physical characteristics, school managements would generally be aware of a student's SC/ST status, even if she were not availing herself of any of these caste-specific benefits. First, last names usually give a good indication of an individual's caste background. Second, even if an individual has dropped her last name to prevent people from inferring her caste, it would be difficult to disguise her SC/ST status since most people know who the "untouchables" are even if the positions of other castes in the local hierarchy are not clear (Srinivas, 1954).

Evidence that this knowledge hurts SC/ST students in terms of school admissions in India is largely anecdotal. The PROBE report(1999), for example, cites instances of teachers' barring SC/ST children from their classrooms, and similar occurrences have been documented by the news media (e.g. Sainath (1999), Behal (2002)). One informal survey found villages in which Scheduled Caste students were discouraged from attending better quality government primary schools outside of their neighborhood and told to attend the closer but poorer quality schools, even though govern-

[4]*Gautama Dharma Sutra* Chapter 12, verses 4 - 6.

[5]*Manu Smriti* Chapter 4, verse 108.

[6]The Karnataka Scheduled Castes, Scheduled Tribes and Other Backward Classes (Reservation of Seats in Educational Institutions and of Appointment or Posts in the Services under the State) Act of 1994.

ment schools are supposed to admit all students who apply regardless of their place of residence (Ramachandran, 2002). With a full sample of secondary schools, this paper can establish whether these instances of caste bias in admissions are the exceptions or the rule in Karnataka.

2.2 Secondary education in Karnataka

Two features of the secondary education environment in Karnataka make it a suitable setting for studying caste bias in admissions - namely the lack of residence requirements in government-run schools and the relatively high share of schools run by private managements.[7] In the United States, the extent to which individuals sort across neighborhoods in some sense limits the degree of sorting across schools since most schools require their students to live within their administratively defined school-attendance area. In Karnataka, however, any type of clustering in schools- by test-taking ability or by caste - should reflect choices about schools and should not be obscured by choices about neighborhoods. Moreover, low internal migration rates in Karnataka suggest that choices about neighborhoods are rarely an issue in any case.[8]

With respect to their managements, schools in Karnataka (and in the rest of India) can be classified as *government*, *aided*, or *unaided* schools, which cater to 48, 35, and 17% of students in 10^{th} grade 2004, respectively. In government schools, the state government appoints teachers, constructs the building of the school, and pays for teachers' salaries and for some infrastructure. In practice, however, local government bodies and voluntary organizations must step in to fund infrastructure such as furniture, toilets, libraries, and drinking water facilities. These schools are supposed to admit any student who applies and waive tuition for SC/ST students.

The managements of aided schools appoint their own teachers but they can still rely on the state to pay for their salaries and to fund a small portion of their maintenance costs. All other expenses such as construction of the school building or investment in basic infrastructure must be met by local government bodies and voluntary organizations. These schools do not have to admit anyone who applies and can exercise choice over the composition of their students; considered to be of better quality than government schools, they often administer entrance exams and conduct interviews to

[7] According to the Department of Public Instruction, 65% of schools in 2002 were run by private managements. The corresponding figure in the KSEEB data was 54%.

[8] In 1991, 70% of individuals in the Census had been born in the village in which they were being interviewed. 88% had been born in the district in which they were being interviewed (There are 29,731 villages with an average area of 6 square kilometers or 2.3 square miles. In 1991, Karnataka was divided into 20 districts with an average area of 10,174 square kilometers or 3,927 square miles.)

ration their slots and maintain long waiting lists of applicants who were not offered admission. Their fees tend to be much higher than the government school rate, but they must still waive these for SC/ST students.[9]

Unaided schools also enjoy complete autonomy over their choice of both teachers and students and maintain long waiting lists, but they receive no funding from the state. Moreover, they are not bound to charge SC/ST and non-SC/ST students different fees.

Schools in Karnataka also differ with respect to their medium of instruction. For the most part, secondary education proceeds in two main languages of instruction - English and the state language Kannada. Since 86% of secondary schools that offer 10^{th} grade instruction begin in grade 8, the allocation of students to schools generally occurs 3 years prior to the matriculation exam, whereas the allocation of students to a certain medium of instruction should occur much earlier. Although there are no regulations barring students from switching mediums before entering 8^{th} grade, it would be difficult after seven years of schooling in another language, and according to one former District Director of Public Instruction, it is not very common (Karanth, 2006). Thus, it is likely that English and Kannada secondary schools operate in separate education markets, and individuals would sort across schools only within their chosen medium of instruction.

Regardless of their school's management type or medium, in order to obtain a matriculation (school-leaving) certificate, students must take an exam that KSEEB administers every year in March, which tests material from seven subjects (three languages, Social Science, Mathematics, Physics/Chemistry, and Biology) from a syllabus that all secondary schools must follow in order to be recognized by Karnataka's Department of Public Instruction. An individual must pass this exam in order to progress to 11^{th} grade and to secure employment in more desirable blue-collar jobs.[10] Privately managed junior colleges (schools that offer instruction in 11^{th} and 12^{th} grade) also use the scores on this exam as admissions thresholds. Appendix A contains excerpts from the 2005 English, Mathematics, and Social Science exams.[11] In order to answer these questions correctly, students clearly need access to schools that teach the entire syllabus, drill students on a regular basis, and offer instruction on answering the

[9] It is very difficult to observe school fees in this setting. Although Karnataka prohibits the collection of capitation fees and gives the state the authority to regulate tuition fees (Governement of Karnataka, 2004), aided and unaided schools can get around this by asking students for "donations" prior to admission, so any schedule of fees would be right-censored. In a survey on household expenditure on education, Tilak (2002) finds that parents sending their children to aided and unaided schools in Karnataka pay more than double the fees than what parents pay when they send their children to government schools.

[10] The matriculation certificate could mean the difference between working as a janitor or working as a driver or secretary.

[11] Aside from the language sections, the Kannada and English versions of the exam are identical.

more subjective questions. Finally, the ranking of all schools by their average performance appears as a list in the newspaper when the results of the exam are announced in May.

3 A simple model of sorting and admissions

This section presents a simple model of sorting and admissions that tries to distinguish the patterns that should hold in the data if schools exhibited a (taste-based) bias against SC/STs in their admissions decisions from those that would arise if schools were simply trying to maximize their average performance on the matriculation exam. Although students can, in principle, choose whichever school they want to attend, the final allocation of students to schools will not result solely from decentralized individual decision-making, or self-sorting, as is the case in some of the neighborhood-sorting literature in the United States that focuses on sorting by race across neighborhoods (Becker and Murphy (2000) and Sethi and Somanathan (2004)). The model assumes that schools can also exhibit preferences over the composition of their student body and can use a variety of instruments such as entrance exams and interviews to restrict certain types of students from attending their school. This double-sided decision-making allows schools' admissions policies to serve as a rationing device in the model and captures certain features of the institutional environment relevant to sorting. The widespread use of entrance exams in non-government schools suggests that prices (or fees) are not sufficient to achieve individuals' or schools' desired allocation. Since reservation policies do not constrain secondary schools, there is nothing to prevent caste from playing an implicit role in admissions decisions.

The model first ignores potential caste differences in the demand for school quality but later shows that standard concerns like liquidity constraints or the non-random placement of schools will not affect the ability of the proposed empirical test to *rule out* taste-based discrimination against SC/ST students. The model also first considers the case in which schools can perfectly screen students by their test-score generating potential. An extension then discusses the case in which schools can only imperfectly back out a student's test-score potential prior to admission and castes differ in their unobservable inputs to test-score production. This framework will highlight the difference between taste-based and statistical discrimination.

Abstracting away from potential gender differences in school-choice and from the sorting of students across different management types - complications that will be addressed in the empirical section of the paper, the model ultimately makes empirical predictions about the relative performance of castes *within* schools that will distinguish an allocation characterized by taste-based discrimination against SC/STs in admissions from one with no such discrimination. The simple empirical test that follows from the model turns out to be quite undemanding in terms of the data required on individuals; it only calls for knowledge of an individual's test-score, caste, and which school she attends.

3.1 Endowments and technology

Individuals can be characterized by their test-taking ability prior to 8^{th} grade θ_i , their achievement gains that will be realized during high school h_i, and their caste $c_i \in \{L, H\}$, where L refers to SC/ST individuals and H to everyone else. This θ_i reflects both innate ability and motivation and *acquired* ability - that is, all past human capital investments made in an individual prior to her entering secondary school, with these investments broadly defined to include things such as spillovers from parental human capital and other complementary inputs to formal schooling like after-school tutoring, a quiet neighborhood, or freedom from laborious household duties. The attribute h_i measures how an individual will interact with her peers and teachers in the classroom and the effects of complementary inputs that might matter more during high school than during elementary school such as exam-specific tutoring or the presence of a parent who has completed secondary school or college. A measure of an individual's overall test-score potential z_i is simply of function of θ_i and h_i. Within each caste-group, z can be represented by distribution functions $F^H(z)$ and $F^L(z)$ with supports $[z^L_{min}, z^L_{max}]$ and $[z^H_{min}, z^H_{max}]$. Initially, assume that no one is liquidity constrained but that $z^L_{min} < z^H_{min} < z^L_{max} < z^H_{max}$ and for any $z \in [z^L_{min}, z^H_{max}]$, $F^L(z) > F^H(z)$. To the extent that parental human capital and asset ownership are correlated with an individual's test-taking ability and achievement gains during high school, this assumption that the non-SC/ST distribution of z dominates the SC/ST distribution is in line with the Census data on caste differences in school completion and asset-ownership presented earlier. This SC/ST disadvantage in terms of parental human capital and income is assumed to outweigh any selection effects stemming from the lower SC/ST secondary school participation rate.[12]

A school j displays an attribute q_j that characterizes its inputs into test-score production such as

[12]According to the 2001 Census, 58% of SC/ST 14 year olds in Karnataka are still in school, whereas the corresponding figure for the non-SC/ST population is 70%.

teacher quality, infrastructure, or pedagogical methods.[13] For simplicity, assume all schools are the same size $\underline{n}$, which represents the optimal number of students required to make a school viable given the costs of constructing and operating a school.[14]

Individuals and schools operate in an area defined by the size of its test-taking population N and its share of SC/ST students λ. The fixed school size assumption implies that an area with with population N supports a number of schools equal to $N/\underline{n}$. Thus, if an area's test-taking population grows by $\underline{n}$, another school can enter to accommodate those students.[15]

An individual's score or success on an exam g is a function of both her own test-score potential and the inputs of the school she attends, and this test-score production function f is monotonically increasing in z, and it is assumed that students with lower test-score potential do not benefit more from school inputs:

$$g_{ij} = f(z_i, q_j) \quad \text{with} \quad \frac{\partial f}{\partial z} > 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial q \partial z} \geq 0.$$

Thus, the theoretical (and eventually empirical) model assumes that, within a school, a student with lower test-score potential cannot attain the same score as a student with higher potential.

Another implication of these assumptions is the irrelevance of caste in test-score production: two individuals of different castes with the same test-score potential when they enter a specific school should achieve the same test-score when they finish:

$$f(z, q \mid H) = f(z, q \mid L).$$

3.2 Decision-making

3.2.1 With no discrimination

Suppose first that individuals and schools do not care about the caste-composition of schools. In choosing a school, individuals only seek to maximize their test-score g_{ij}. Since this test-score is a function of their own test-score potential z_i and school inputs q_j, this amounts to seeking admission

[13] To the extent that schools have a history of admitting students of particular test-taking ability levels, peer-effects can also be subsumed in q_j.

[14] As will become clear, this fixed school-size assumption is not critical for the qualitative results of the model. It simply eliminates the decision that schools need to make about capacity. For the purposes of this model, this fixed capacity could be considered to be school-specific as well.

[15] This accords with at least what is written in the DPI's application to start a private school (DPI, 2004). Potential school managements must furnish information about the latest Census population counts from their proposed school site in addition to the number of students attending each year of primary school (grades 1 to 7). The DPI, however, does not claim to use any specific population-thresholds when granting permission to open a school.

at schools offering the highest levels of q_j. Thus, as will be evident later, another critical assumption of the model is that an individual i's utility from school j, U_{ij}, is an increasing function of q_j:

$$U_{ij} = u(g_{ij}) = u(q_j) \quad \text{with} \quad \frac{\partial u}{\partial q_j} > 0.$$

A school management sets its admissions requirements to maximize its average test-score $\tilde{g}_j$, and to satisfy the school-size constraint imposed by $\underline{n}$. Given that test-scores are a function of z and q, this amounts to maximizing the average level of test-score potential in the school $\tilde{z}_j$. Thus, it must choose a threshold level of test-score potential $\underline{z}_j$ and deny admission to applicants whose levels of z_i lie below it:

$$\max_{\underline{z}_j} \quad W_j = w(\tilde{g}_j) = w(\tilde{z}_j)$$
$$\text{s.t.} \quad n_j = \underline{n}$$

Underlying this objective could be cost minimization (individuals with higher z_i might be easier to teach) or pure reputation concerns (to the extent that reputation is a function of the publicly observable average school performance on the state-wide exam). This objective could be achieved by setting fees, with the standard assumption in the sorting literature that individuals with higher test-score potential are willing to pay more for school quality ($\frac{\partial^2 u}{\partial q \partial z} > 0$), or by providing a menu of school inputs that are particularly attractive to students of high test-taking ability (e.g. teachers with a particular education background). Assume, however, that schools can more directly screen students by observing z_i (through interviews or entrance exams) and admitting only the top $\underline{n}$ students of their applicant pool.

3.2.2 With discrimination

Now suppose that caste also enters the admissions process. If, in addition to maximizing their average test-score potential, schools also take their caste-composition into consideration and try to limit their share of SC/ST students λ_j, then they potentially must choose *caste-specific* thresholds and their objective function becomes:

$$\max_{\{\underline{z}_j^H, \underline{z}_j^L\}} \quad V_j = v(\tilde{z}_j, \lambda_j).$$
$$\text{s.t.} \quad n_j^H + n_j^L = \underline{n}$$

with

$$\frac{\partial v}{\partial \lambda_j} < 0$$

This preference over school caste-composition could stem from inherent prejudice on the part of school managements, from pressure from the more dominant non-SC/ST population who might want their children schooled separately from SC/ST students, or from fear of being "outcasted" (Akerlof, 1976) for mixing castes. Thus, relative to the case in which they only cared about maximizing $\tilde{z}_j$, schools now may be willing to make a trade-off at the margin: to sacrifice some SC/ST students to admit non-SC/ST students of lower demonstrated z_i in order to keep λ_j below a certain level. This notion of discrimination accords with Becker's (1957) standard that a discriminatory firm must forfeit profits if it exhibits a taste for discrimination. Again, suppose schools choose students by directly observing test-score potential z_i and caste, admitting n^L of the top SC/ST applicants, and offering the remaining $\underline{n} - n^L$ slots only to its best non-SC/ST applicants.[16] If the marginal cost of giving up some average test-score $\tilde{z}_j$ does not always outweigh the marginal benefit of decreasing the SC/ST school share λ_j, then all schools cannot maximize their objective function V_j simply by maximizing their average test-score potential $\tilde{z}_j$.

3.3 Equilibrium allocations

The rest of this section uses a stylized four-school example to illustrate the equilibrium patterns that should hold in the middle of the school performance distribution and in the tails for the two cases described above - when the allocation of students is completely determined by their test-taking score potential and when an additional concern for caste composition guides sorting.

3.3.1 With no discrimination

When individuals do not care about the caste-composition of their schools and schools only set out to maximize W_j, individuals with high test-score potential attend schools with high levels of school quality. The school that offers the highest level of inputs will admit the top $\underline{n}$ students in the area in terms of z. The school ranked immediately below it in terms of q_j will select the top $\underline{n}$ individuals from the remaining $N - \underline{n}$ students in the population. Thus, in equilibrium, schools will be defined

[16] As in the non-discriminatory case, schools could also accomplish this indirectly through prices since non-SC/ST individuals might also care about the λ_j of the school they attend and therefore may be willing to pay a premium for going to school with few SC/ST classmates or schools could offer amenities that might be particularly attractive to non-SC/ST students such as non-SC/ST teachers or a school situated far from a water source used by the area's SC/ST population.

by intervals $(\underline{z}_j, \overline{z}_j)$, and all students within a school j will display a level of test-score potential that falls within this interval regardless of their caste. That is,

$$\underline{z}_j^H = \underline{z}_j^L$$

for all schools except the worst school which admits the SC/ST student displaying lowest level of test-taking ability z_{min}^L, and

$$\overline{z}_j^H = \overline{z}_j^L$$

for all schools except the best school which accepts the non-SC/ST student with the highest realization of ability z_{max}^H. The ranking of schools in terms of their average performance $\tilde{g}_j$ will mirror the ranking of q_j, since the schools with the highest school quality will attract the students with the highest test-score potential.

Figure 2 shows how equilibrium sorting divides $F^H(z)$ and $F^L(z)$ into schools of equal size when $\lambda = 0.20$ and $q_j > q_{j+1}$ for $j \in \{1,2,3\}$. To simplify presentation, z appears uniformly distributed within each caste-group, with supports $[z_{min}^H, z_{max}^H] = [0.2, 1]$ and $[z_{min}^L, z_{max}^L] = [0, 0.8]$.[17] Individuals whose z_i lies in the interval $[\underline{z}_1, z_{max}^H]$ attend School 1, those with $z \in [\underline{z}_2, \underline{z}_1)$ and $z \in [\underline{z}_3, \underline{z}_2)$ are matched to Schools 2 and 3, respectively, and finally those with $z \in [z_{min}^L, \underline{z}_3)$ have to attend the worst school in terms of q_j, School 4.[18]

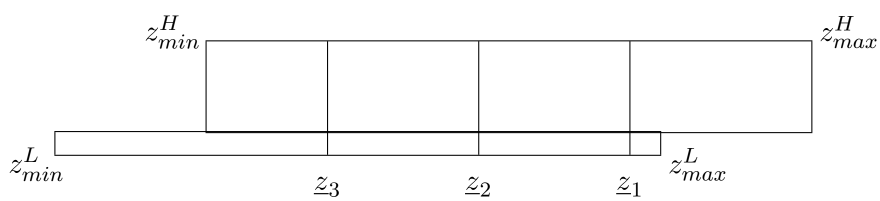


Figure 2
Allocation with no discrimination, 4-school case

In this allocation no individual can access a higher level of q_j and no school has the opportunity to increase its average test-score by altering its admission-thresholds. An individual with $z_i < \underline{z}_1$, for

[17] The importance of distributional assumptions for the *empirical* analysis will be discussed in the section that summarizes the model's empirical predictions.

[18] Since schools in the basic model are all the same size, these thresholds solve the following:

$$\lambda[1 - F^L(\underline{z}_1)] + (1-\lambda)[1 - F^H(\underline{z}_1)] = \tfrac{1}{4} \quad , \quad \lambda[F^L(\underline{z}_1) - F^L(\underline{z}_2)] + (1-\lambda)[F^H(\underline{z}_1) - F^H(\underline{z}_2)] = \tfrac{1}{4}$$

$$\lambda[F^L(\underline{z}_2) - F^L(\underline{z}_3)] + (1-\lambda)[F^H(\underline{z}_2) - F^H(\underline{z}_3)] = \tfrac{1}{4} \quad , \quad \lambda[F^L(\underline{z}_3)] + (1-\lambda)[F^H(\underline{z}_3)] = \tfrac{1}{4}$$

example, cannot switch to School 1 and access the highest level of school inputs q_1. School 1 would deny her admission since she would have to replace a student whose test-score potential does lie in the $(\underline{z}_1, z^H_{max})$ interval - an exchange that would only decrease its level of average test-score potential, $\tilde{z}_1$. No school can improve its average level of test-score potential by increasing or decreasing its admissions thresholds. An increase of its admissions cutoff $\underline{z}_j$ drops school size below $\underline{n}$; a decrease would only lower average test-score potential $\tilde{z}_j$.[19]

What does this type of equilibrium sorting imply for the allocation of SC/ST students across schools? Will SC/ST students disproportionately populate the worst performing schools? Although the probability that SC/ST and non-SC/ST students attend a school j should be the same *conditional* on z_i, the unconditional probability of going to the best performing school should be smaller for SC/ST students since $F^H(z)$ dominates $F^L(z)$. In the above example, when students are perfectly sorted across schools by their test-score potential, only 5% of all SC/ST students attend School 1 - the school with the highest average performance $\tilde{g}_1$, whereas 45% end up in the worst performing school - School 4. This type of selection into schools is considerably less skewed for the non-SC/ST student population: 30% attend School 1 and 20% School 4. Relative to their population share of 20%, this leads to under-representation of SC/ST students in the best school (4%) and over-representation in the worst school (36%).[20] Thus, a concentration of SC/ST students in the lowest performing schools can occur even when caste does not figure into the decision-making of individuals and schools.

This type of allocation, however, has an additional implication for the relative performance of castes *within* schools. Because the cutoffs $\underline{z}_j$ and $\overline{z}_j$ for each caste are identical in the schools that fall within the common support of the two ability distributions, in terms of performance, the marginal SC/ST student should appear identical to the marginal non-SC/ST student in her school in all schools except the worst school, since

$$f(\underline{z}_j, q_j|L) = f(\underline{z}_j, q_j|H).$$

[19]Note that such an equilibrium can also be supported by prices $p^\star$ when $\frac{\partial^2 u}{\partial q \partial z} > 0$. In the 4 school example, a school j would have to set its fees $p^\star_j$ so that $U_{ij} = U_{ij+1}$ when $z_i = \underline{z}_j$. This would guarantee that for inframarginal students (those students with $z_i \neq \underline{z}_j$ for any j), $U_{ij} > U_{ik}$ for $j \neq k$ when $z_i \in (\underline{z}_j, \overline{z}_j)$. Schools would also have no incentive to alter their fees. Increasing them would lead to a violation of the minimum school-size constraint. If p_j were set below $p^\star_j$, individuals with $z_i < \underline{z}_j$ could attend school j. Because the willingness to pay for q_j increases with z_i, these individuals would be outbid by individuals with $z_i \in (\underline{z}_j, \overline{z}_j)$ until $p_j = p^\star_j$.

[20]These shares have been calculated from the following expressions:

$$\lambda_1 = 4\lambda[1 - F^L(\underline{z}_1)] = (0.8)(0.05) = 0.04$$

$$\lambda_4 = 4\lambda[F^L(\underline{z}_3)] = (0.8)(0.45) = 0.36$$

Similarly, the best students of each caste should achieve the same scores in all schools except the best school. Note that although the dominance assumption (i.e. $F^H(z)$ dominates $F^L(z)$) helps to explain why SC/ST students are more likely to attend the worst performing schools, it is not necessary for deriving these within-school predictions.

3.3.2 With discrimination

What happens when discriminatory frictions are introduced to the sorting outlined above - specifically, when an additional concern for a student's caste motivates schools' admissions decisions? First, note that at the allocation without discrimination, a school cannot decrease its SC/ST share λ_j without also incurring a loss in test-score potential $\tilde{z}_j$ or violating the minimum school-size constraint. If, however, $|\partial v/\partial \lambda_j| > \partial v/\partial \tilde{z}_j$, a school can benefit from replacing its SC/ST students displaying the lowest levels of z in the school with non-SC/ST students of slightly lower test-score potential since they are willing to sacrifice some $\tilde{z}_j$ in return for a decrease in λ_j. When this occurs, a discriminating school will be defined by *caste-specific* intervals - $(\underline{z}_j^L, \overline{z}_j^L)$ and $(\underline{z}_j^H, \overline{z}_j^H)$ - with $\underline{z}_j^L > \underline{z}_j^H$.

The diagram of Figure 3 corresponds to a situation in which all schools need to make adjustments in their thresholds relative to the non-discriminatory allocation. This could occur when marginal utility from $\tilde{z}_j$ increases with $\tilde{z}_j$ ($\frac{\partial^2 v}{\partial \tilde{z}^2} > 0$) and marginal disutility from λ_j increases with λ_j ($\frac{\partial^2 v}{\partial \lambda^2} < 0$) - that is, drops in $\tilde{z}_j$ are more damaging for the schools serving students with the highest levels of test-score potential, and decreases in λ_j benefit schools with high SC/ST shares more.[21] Thus, if $|\partial v/\partial \lambda_j| > \partial v/\partial \tilde{z}_j$ for School 1 at the non-discriminatory allocation, then the other schools will also benefit from adjusting their thresholds away from the non-discriminatory thresholds since $\tilde{z}_j$ is decreasing and λ_j is increasing along the performance distribution: in terms of forgone utility from $\tilde{z}_j$, these schools lower down the performance distribution lose less than School 1, and in terms of increments to utility from reducing λ_j, they gain more. As in the equilibrium with no discrimination, the process of selecting students is recursive. School 1 sets its admissions thresholds with $\underline{z}_1^H < \underline{z}_1^L$. If School 2 were to set $\underline{z}_2^H = \underline{z}_2^L$, then it would have a higher $\tilde{z}_2$ than in the non-discriminatory allocation because of the SC/ST students shut out of School 1, which should make it less likely to adjust. It would also have a larger λ_2, however, which should make it more likely to adjust. The diagram in

[21]This corresponds to a situation when reputation is more sensitive to movements in the z-composition of their students at high levels of $\tilde{z}_j$ (where schools have more of reputation to protect) and when schools worry more about their caste-composition as their share of SC/ST students increases (when the risk of "tipping over" might be greater or when it is easier to identify SC/ST students as a group).

Figure 4 assumes that the latter effect outweighs the former for Schools 2 and 3, but note that even if School 2 does not adjust its thresholds, School 3 surely would since it would be faced with exactly the same decision problem (applicant pool) as in the pure non-discriminatory case.

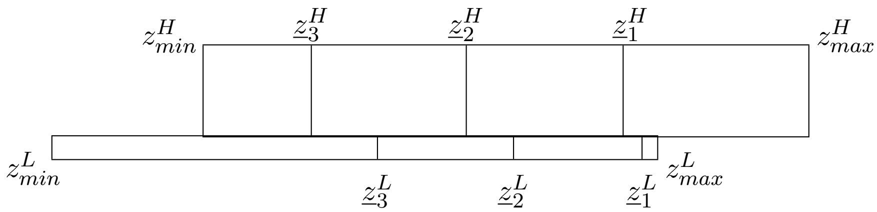


Figure 3
Allocation with discrimination, 4 school case

Although the unconditional probability of attending the best (worst) school is still lower (higher) for SC/ST students, in this case in which caste does matter for school assignment, the conditional probabilities of attending a school j are not equivalent across castes. Now there are mismatches in the sense that a non-SC/ST individual enjoys a higher (lower) probability of attending a high (low) performance school than a SC/ST individual of the same test-score potential. In this particular case, when all schools adjust their thresholds, the marginal SC/ST student should outperform her non-SC/ST counterpart within the school in all schools except the worst school, and the best SC/ST student in the school should achieve a higher score than the best non-SC/ST student in all schools except the best school. In a situation in which all schools are not simultaneously adjusting, discriminatory schools will have $\underline{z}^L_j > \underline{z}^H_j$ and the marginal SC/ST student will outperform her non-SC/ST classmate. Because of the recursive nature of admissions (schools can only select from students who are not attending another school offering a higher level of q_j), schools that follow a discriminatory school in the ranking of q_j would also have their best SC/ST students achieving higher scores than their best non-SC/ST students.

3.4 Comparative statics

Before extending the model further to consider caste-differences in demand and imperfect observability of z, this section briefly outlines the importance of an area's characteristics such as N and λ in detecting taste-based discrimination in admissions. First, recall that increases in N in the model are tantamount to increases in the number of schools. In the extreme case in which N is so low that an area only supports one school, sorting of any kind is just not possible. Differences across castes in observed admissions thresholds within the school would simply reflect the difference in the distributions $F^H(z)$ and $F^L(z)$.[22] However, as more and more schools divide an area's overall performance distribution into (possibly caste-specific) intervals of z, differences across castes within schools should be less contaminated by differences that hold in the population. Thus, in order to disentangle differences in admissions thresholds that arise from discrimination from the downward bias associated with the dominance of $F^H(z)$ over $F^L(z)$ in the population, an area must contain a sufficient number of schools.

An area's overall SC/ST share λ also determines the extent to which the discriminatory and non-discriminatory allocations can be distinguished. If the convexity of V_j with respect to $\tilde{z}_j$ and the concavity with respect to λ_j remain constant across areas (i.e. if the tradeoff between average test-score potential and caste composition remains the same), then in the context of a very low λ, there may be no school for which the marginal gain from decreasing its SC/ST composition outweighs the marginal loss in average test-score potential at the non-discriminatory allocation, and thus there may be no adjustments in thresholds even when λ_j enters schools' utility functions.

In the reference non-discriminatory allocation, partial changes in λ serve to shift the z-thresholds that define schools to the left because of the fixed school-size constraint and the assumption that the non-SC/ST distribution of test-score potential dominates the SC/ST distribution. In the case in which the first school contains both SC/ST and non-SC/ST students (as in the 4 school example above), a partial increase in λ gives the average SC/ST applicant's test-score potential more relevance in the setting of a threshold that would satisfy the school-size constraint, and because the average SC/ST applicant displays a lower level of z given the dominance assumption, the threshold must shift to the left.[23] Because of the recursive nature of admission (School j accepts students that School $j-1$ does not admit), the thresholds of all other schools would also shift to the left. In the non-discriminatory allocation, this shift in thresholds does not effect the equilibrium in a qualitative way: schools still set identical admissions thresholds for each caste

This is not the case when schools discriminate, in which case partial changes in λ actually induce

[22]It is possible, of course, that managements of singleton schools could discriminate and keep some SC/ST students out of secondary school entirely, and it could be these "monopolist" schools that would be the most likely to discriminate in the absence of any competitive pressure (Becker, 1957). In an environment in which secondary school participation is not universal, however, it would be difficult to separate the effects of discrimination and different selective pressures on SC/ST and non-SC/ST individuals in these single-school settings.

[23]That is, in the expression to determine School 1's threshold - $\lambda[1-F^L(\underline{z}_1)]+(1-\lambda)[1-F^H(\underline{z}_1)]=\frac{1}{4}$ - more weight will be placed on $F^L(z)$ relative to $F^H(z)$.

a behavioral effect. When the z-thresholds that define schools in the non-discriminatory allocation shift to the left, $\tilde{z}_j$ decreases in each school, while λ_j increases. Given the convexity of V_j with respect to $\tilde{z}_j$ and the concavity with respect to λ_j, this shift should make discriminatory schools more likely to adjust and increase the probability of finding mismatches in admissions thresholds within schools. Thus, caste differences in admissions requirements should be more evident in areas with a sufficient share of SC/STs in the population.

3.5 Caste-differences in demand

The analysis thus far has ignored potential caste differences in the demand for school quality q_j and has implicitly assumed that conditional on test-score potential z, SC/ST and non-SC/ST students display the same willingness or ability to pay for q. What if this were not the case? Such a situation could arise because of SC/ST-specific liquidity constraints that do not affect the production of z but do influence an individual's choice of a particular school or because of the non-random placement of secondary schools in non-SC/ST predominant areas and the consequent higher transportation costs for SC/ST students.[24] If such differences mattered, then SC/ST students would be shut out of schools that non-SC/ST students of slightly lower z could afford and would have to attend lower quality schools with non-SC/ST students demonstrating lower levels of z - an allocation that mirrors the one that results when schools actively try to limit their SC/ST composition. These caste-differences in demand could therefore generate an allocation that appears discriminatory even in the absence of any prejudice on the part of schools against SC/ST students. They cannot, however, obscure prejudice if it is present since they only reinforce the discriminatory patterns that would hold in the data. That is, they cannot generate an allocation that appears non-discriminatory when schools do in fact discriminate against SC/STs. Thus, a necessary (although not sufficient) condition for taste-discrimination against SC/ST students in admissions is an SC/ST performance advantage when comparing the marginal students in a school since $\underline{z}_j^L > \underline{z}_j^H$ implies $\underline{g}_j^L > \underline{g}_j^H$. Thus, although it would be difficult to establish that schools do in fact exhibit a taste for discrimination, it is relatively easy to prove that they do *not*.

[24] In what will later be defined as a school-attendance area, the correlation between the number of English schools in an area and the SC/ST share of the entire population is around -0.13, and the corresponding correlation for Kannada schools and an area's SC/ST share is -0.18.

3.6 Imperfect screening and caste-specific inputs

At this point, a skeptic might argue that the stricter admissions criteria for SC/STs would not translate into a performance advantage within schools as predicted by the model because of imperfect observability of z and caste-specific inputs that might enter test-score production. Although SC/STs might enter a school as a more select sample in terms of their test-taking ability demonstrated prior to 8^{th} grade, differential treatment within the classroom post-admission, differential access to complementary inputs that matter more in secondary school than in primary school (e.g. after-school tutoring or an educated parent), or differences in motivation because of caste-specific labor market returns to matriculation could hinder their achievement gains during high school and make the marginal SC/ST student appear equal to or worse than the marginal non-SC/ST student at the time of the exam . This differential treatment could range from outright harassment from teachers or students to subtler forms of discouragement such as low-expectations.[25]

Although such skepticism is warranted since these kinds of caste-specific inputs likely play a role in test-score production, it is misguided in the sense that it entails assuming schools simultaneously display taste-based discrimination against SC/ST students in admissions and overestimate their test-score potential. To see this, suppose schools cannot perfectly screen students with respect to z and can only rely on their test-taking ability demonstrated prior to 8^{th} grade θ (which can be considered the raw score on a school's entrance exam) and some estimate of the gains h they will make during high school that varies systematically by caste. If caste-specific inputs really mattered, then over the years each school should have formed some expectation about the difference between h_j^H and h_j^L in their applicant pool and a method for translating the θ_j^H and θ_j^L of their applicants into test-score potential by appropriately discounting θ_j^L. Figure 4 shows that when $h_j^H > h_j^L$, schools must set a higher cutoff in terms of θ for their SC/ST applicants if they want to set $\underline{z}_j^H = \underline{z}_j^L$.[26] The line labeled Z depicts the line of equality in terms of test-score potential for different levels of h and θ. The distance between $\underline{\theta}_j^H$ and $\underline{\theta}_j^L$ represents the amount of *statistical* discrimination a school must employ against an SC/ST student in order to maximize its average test-score. Schools set higher admissions standards for SC/STs in terms of θ only so that at the time of the exam, SC/STs and non-SC/STs are

[25]A recent World Bank study from a northern Indian state (Hoff and Pandey (2004)), for example, found no significant difference between "low-caste" and "high-caste" boys' ability to solve mazes when their identity was not made salient. In their treatment group, however, when the name and caste of each child was announced prior to the commencement of the game, a significant caste-gap emerged and "low-castes" solved 20% fewer mazes, a result they attribute to the anticipation of impartial treatment and the resulting decline in motivation.

[26]Glenn Loury kindly suggested demonstrating this point with this type of graph.

identical in what matters for final outcomes - test-score potential z.

A school that engages in taste-based discrimination, however, would set a higher SC/ST threshold in terms of final outcomes, $\underline{z}_j^H > \underline{z}_j^L$. The would hold SC/ST students to the higher standard represented by the line Z^* by admitting only those with $\theta > \theta_j^{L^*}$. In this case, SC/STs would not outperform their non-SC/ST classmates only if schools somehow ignored the SC/ST disadvantage with respect to h and failed to discount SC/ST students' θ appropriately. The distance between $\underline{\theta}_j^L$ and $\underline{\theta}_j^{L^*}$ represents the amount of "statistical discrimination" *in favor* of SC/STs a taste-based discriminating school must employ in order to equate the performance of the marginal SC/ST and non-SC/ST students in their school. It is unlikely, however, that schools would both exhibit prejudice against SC/STs and at the same time overestimate their test-score potential. Thus, if the marginal SC/ST student does not outperform her non-SC/ST counterpart, then her school does not engage in discriminatory admissions.

3.7 Empirical predictions

These predictions about relative admissions thresholds translate into very simple empirical specifications that only require knowledge of a test-taker's caste and the identity and location of her school. In both the non-discriminatory and discriminatory allocations, the coefficient on an SC/ST indicator in a simple ordinary least-squares regression of test-performance on caste,

$$g_i = \alpha_0 + \alpha_1 SCST_i + \epsilon_i, \quad (1)$$

should be negative since it combines the effects of caste-discrepancies with respect to test-score potential z and to school quality q. SC/ST students should perform poorly relative to non-SC/ST students in the population because they demonstrate lower levels of test-score potential on average (possibly because of both lower levels of test-taking ability acquired prior to 8^{th} grade θ and smaller gains made throughout high school h) and because they have to attend schools offering low amounts of q_j.

Although the qualitative predictions of the model did not require any assumptions on the two distributions of test-taking ability, measuring differences in thresholds ($\underline{z}_j^L - \underline{z}_j^H$ and $\overline{z}_j^L - \overline{z}_j^H$) will implicitly entail making such assumptions. As Figure 3 demonstrates, if z were uniformly distributed within each caste-group, then a comparison of means ($\tilde{z}_j^L - \tilde{z}_j^H$) in schools that fall entirely within the common support of the two ability distributions would be sufficient to detect any differences in admission thresholds. If schools exhibited a taste for discrimination against SC/STs, this difference in means would be positive. That is, in a regression like (1) that also included a set of school dummies, δ_j,

$$g_{ij} = \beta_0 + \beta_1 SCST_i + \delta_j + u_{ij}, \tag{2}$$

the coefficient on SC/ST_i would measure the average within-school caste-gap ($\tilde{z}_j^L - \tilde{z}_j^H$) and would be positive if schools' admission standards were prejudiced against SC/STs. More specifically, it would be positive in the middle of the distribution of school performance. Although SC/ST students could score better or worse on average relative to the non-SC/ST students attending the same school in the tails of a school-attendance area's performance distribution depending on the relative positions of the distributions' supports, they should perform better at least in the sections of the two distributions that completely overlap if schools were discriminating. Thus, if schools are assigned quintiles Q based on the distribution of average performance of schools in their school-attendance area, in a regression that also includes an indicator for the position of an individual's school in the area's performance distribution, δ_{Qj},

$$g_{ij} = \beta_0 + \beta_1 SCST_i + \beta_Q SCST_i * \delta_{Qj} + \delta_j + u_{ij}, \tag{3}$$

the sum $\beta_1 + \beta_Q$ should be positive in the middle quintiles.[27] Given that an area with mutually exclusive quintiles contains at least five schools, there should be less of a concern that the lack of sorting opportunities associated with low levels of N would be biasing the estimated differences in thresholds of Equation (3) towards the difference in the overall means of $F^H(z)$ and $F^L(z)$.

It is easy to imagine, however, other shapes of distributions in which the SC/ST mean in some intervals would still be lower than the non-SC/ST mean despite a higher threshold $\underline{z}_j^L$ for SC/ST students. Table 2 presents simulation results from two normal distributions with unit variance. In each of 10,000 replications, 800 observations were drawn from $F^H(z)$ and 200 from $F^L(z)$, and quartiles of the resulting overall distribution correspond to the schools of Figures 2 and 3, with the fourth quartile corresponding to School 1, the best school. Columns 1 - 4 demonstrate that as long as $F^H(z)$ dominates $F^L(z)$, SC/STs will display a lower within-school mean despite identical admissions thresholds. There is a very high likelihood that $\tilde{z}_j^H > \tilde{z}_j^L$ even in the schools in the middle of the distribution when $F^H(z)$ is shifted to the right of $F^L(z)$. When, for example, the SC/ST distribution has a mean of zero and the non-SC/ST distribution a mean of 0.5, non-SC/STs outperform their SC/ST classmates 72% and 74% of the time in the schools in the second and third quartiles, respectively. Thus,

[27] Without knowledge of the exact distributions $F^H(z)$ and $F^L(z)$ and the precise relationship between $\frac{\partial^2 v}{\partial z^2}$ and $\frac{\partial^2 v}{\partial \lambda^2}$, it is not possible to make any predictions on how the magnitude of this SC/ST advantage should change across quintiles.

even if schools were to deviate slightly from identical thresholds and set $\underline{z}_j^L > \underline{z}_j^H$, SC/ST students might still perform worse *on average* than their non-SC/ST classmates when $F^H(z)$ dominates $F^L(z)$ despite facing stricter admissions standards. In this case, the OLS school fixed-effects regressions, which estimate differences in the conditional means of the two distributions, could underestimate differences in admissions thresholds and obscure evidence of prejudice in admissions. Columns 5-8 also demonstrate that focusing on the worst performing student from each caste-group also has its problems because SC/STs are drawn from a distribution with fewer observations. Even when $F^H(z)$ and $F^L(z)$ are identical, there is an 80% chance that the worst non-SC/ST student displays lower z than the worst SC/ST student in the school. Likewise, Columns 9-12 show that there is an 80% chance that the best non-SC/ST student displays a higher z.

Quantile regression (Koenker and Basset, 1978) offers a way to get around these issues by estimating differences in conditional quantiles - or differences at other points of the SC/ST and non-SC/ST distributions within schools. The upper and lower quantiles are presumably closer than the mean to each group's admission thresholds and would be less influenced by the overall shapes of the distributions within schools; they also would be less plagued by the law of large numbers that makes the worst-performing SC/ST student appear better than her non-SC/ST counterpart (and the best-performing SC/ST student appear worse than her non-SC/ST counterpart) within the school even when admissions thresholds are identical. Unfortunately, solving the optimization problem associated with a quantile version of Equation (2) is not so straightforward, given the number of fixed effects associated with δ_j.[28]. In the OLS case, when the object of interest being estimated is $\beta_1 = E(g|SCST = 1) - E(g|SCST = 0)$, transforming the data into deviations from means dramatically reduces the dimensionality of the estimation problem. The fixed effect essentially drops out of the transformed data since the expectations operator is linear. In the case of quantile regression, however, such a transformation is not possible; quantiles are not linear operators and fixed effects cannot be so easily purged from the problem.

To maintain the spirit of quantile regression with fixed effects in the most straightforward way, this paper proposes two methods. The first directly calculates the percentiles of the SC/ST and non-SC/ST distributions of test-scores in each school and simply estimates their difference for all schools and for schools in each quintile of a school-attendance area's performance distribution. That is, it

[28]Even if the empirical analysis were limited to a single year, there could be up to 8,765 indicators associated with δ_j. In the full sample, in which δ_j represents school-year fixed effects, this number skyrockets to 61,591.

estimates for each percentile p the mean of

$$g_j^p(L)_Q - g_j^p(H)_Q \quad \text{for} \quad p = 0.10, 0.20, ..., 0.90 \ \text{ and } \ Q = 1, ..., 5.$$

The second method involves standardizing schools' overall distributions so that they all have the same test-score values associated with each percentile and running quantile versions of Equations (2) and (3) on the rescaled data. The overall distribution of *percent* of all test-takers can serve as the "numeraire" distribution, and each test-taker can be assigned a percentile (0.05, 0.10, ..., 0.95, 1.00) within her school and matched to the midpoint between the corresponding percentile in the numeraire distribution and the next lower percentile. If, for example, a student's score put her in the 85^{th} percentile of her school, she would be assigned a value of g equal to $\frac{g_{0.80}^{numeraire}+g_{0.85}^{numeraire}}{2}$. Thus, in a regression like

$$g_{ij}^{rescaled} = \beta_0 + \beta^{quantile} SCST_i + u_{ij}, \tag{4}$$

$\beta^{quantile}$ would measure the average difference between each quantile of the SC/ST and non-SC/ST distributions within schools. Since the δ_j in the OLS framework aimed to capture pure location shifts of the distribution of g_i associated with each school, standardizing the data in this way obviates the need for estimating a separate fixed effect for each school. Schools with less than 20 people (i.e. schools in which it was not possible to define mutually exclusive 0.05-percentiles), however, cannot included in such an exercise.

Although restricting attention to areas in which quintiles can be defined guarantees a sufficient number of schools for sorting, a more explicit look at the relationship between N and threshold differences would allay concerns that an SC/ST disadvantage in the population was creeping into schools and obscuring evidence of discrimination. Similarly, since increases in an area's SC/ST share λ also increase the likelihood that discriminating schools would benefit from setting caste-specific thresholds, a direct examination of the effects of λ on differences across castes within schools would further distinguish the non-discriminatory and discriminatory equilibria. Thus, in a regression that uses the school-level differences in caste-specific percentiles described above as its outcomes of interest,

$$g_j^p(L) - g_j^p(H) = \gamma_0 + \gamma_1 N + \gamma_2 \lambda + \xi_j, \tag{5}$$

where area-subscripts on N and λ have been suppressed for clarity, the magnitude of γ_1 demonstrates the extent to which low opportunities to sort bias differences in thresholds downwards, and the magnitude of γ_2 gives an indication of the extent to which low SC/ST shares in areas preclude schools with discriminatory preferences from discriminating.

4 Empirical Analysis

4.1 Data and descriptive statistics

The empirical analysis tests the model's predictions about the relative performance of castes within schools using (i) administrative data from the universe of test-takers of the 10^{th} grade matriculation exam administered by the Karnataka Secondary Education Examination Board in March from 1996-2004, (ii) geographic information on the location of schools, and (iii) village-level population data from the Census of India. The following sections describe the data and discuss the creation of the samples and school-attendance areas used in the empirical analysis.

4.1.1 Test-score sample

Each yearly data file from KSEEB provides information on caste, gender, exam experience (i.e. repeat-status), and age for all test-takers in addition to their raw scores and pass-status for all 7 subjects of the matriculation exam. These files also link individuals to schools. On the exam booklet, a student identifies herself only with a pre-assigned registration number, and graders can see no additional information. KSEEB, however, knows her demographic characteristics from an application that she had to submit with signed consent and verification by her high school's principal.

The Board identifies 5 broad caste-categories: Scheduled Caste, Scheduled Tribe, Other Backward Class (OBC), a General category, and a category for individuals who are not native to Karnataka.[29] The OBC category has always been controversial, especially in Karnataka, and it has often been contended that OBC status does not necessarily reflect backwardness but rather political lobbying power (Srinivas (1968), Jenkins(2003)). Therefore, the empirical analysis of this paper treats caste as a binary category and classifies individuals as SC/ST and non-SC/ST.

Most of the 6,337,079 individuals that took the exam during the 1996-2004 period attended officially recognized schools that teach the KSEEB syllabus and offer 10th grade instruction. The 2.05 % classified as *private candidates*, however, could be students who were not enrolled in any school but who took the exam in order to obtain their matriculation certificate or individuals who pursued

[29]SC/ST and OBC status is determined within a state. That is, an individual may be classified as a scheduled caste in one state but not in another.

their education in unrecognized, informal education institutions. The sample does not include these candidates since they possibly did not attend any school (which makes them irrelevant to any question of sorting across schools) and since this paper is not concerned with how students sort across the formal and informal education sectors. Accordingly, the sample also does not include the 0.05% of students that come from schools sending only 1 student to the exam since information on other students in the grade would be missing and since these singleton students might be misclassified private candidates.

Because the model's predictions apply to students admitted from the same applicant pool and since schools would have admitted students classified as *repeaters* at least a year before first-time test-takers, much of the analysis ignores the 25.54% of test-takers who are repeaters, although, as will be shown, their inclusion changes none of the results. Repeaters, however, perform worse on the exam than first-time test-takers, and SC/ST students demonstrate a higher tendency to take the exam as repeaters.[30] Therefore, this type of exclusion should only bias the results *in favor* of finding caste-discrimination.

The sample is also confined to test-takers who can roughly be characterized as "school-age" - that is, between the ages of 14 and 20. If students enter first grade at the age of 6 in June, then by the time of the 10th grade exam in March, they should be 15 or 16 years old if they never take any time off or repeat a grade. To account for late-starters and minor interruptions, the sample also includes test-takers up to the age of 20, although it is unlikely that students over the age of 18 actually sit in the classroom (Karanth, 2006). Only 3.68% of test-takers in the 2002-2004 period universe of test-takers do not meet this age criterion. SC/ST test-takers, however, are over-represented in the non-school-age category, where they account for 39.30% of the group, compared to their share of 23.62% among those who are school-age. Since older students tend to do worse on the exam, this omission would again bias the results in favor of finding discrimination against SC/STs.

Over 95% of test-takers attend a school that teaches in either English or Kannada, but besides these two main languages of instruction, schools can operate in other languages to accommodate Muslim students seeking instruction in Urdu and various linguistic minorities within the state that constitute majorities in other states. Since universities and most places of employment in both the public and private sectors do not employ these other languages, school quality and the composition

[30] In 2004, for example, around 65% of first-time test-takers passed the exam, whereas only 30% of repeaters did. In 2004, the Board classified around 34% of SC/ST test-takers and 18% of non-SC/ST test-takers as repeaters.

of students probably have little bearing on the decision to attend a linguistic-minority school. This paper is not concerned with the additional sorting generated by schools that cater to the 4.54 % of linguistic minority test-takers, so these schools, which constitute 7.96% of all KSEEB schools, are removed from the sample. The final sample, therefore, includes around 90% of all individuals who took the March exam between 1996 and 2004.

Table 3 describes the characteristics of these first-time test-takers of school-age who took their exams in English or Kannada schools that contained both SC/ST and non-SC/ST students. The variable *pass* measures an individual's overall pass-status for the entire exam, which corresponds to a minimum score of 35% of all possible points (i.e. 219 out of a total of 625 points) and a minimum of 30 points in each of the seven subjects of the exam, while *percent* represents the percent of all possible points earned on the exam. Although *percent* corresponds to the continuous distribution of test-taking potential assumed by the model, the variable *pass* suggests the magnitude of caste-differences with respect to later-life outcomes since an individual must pass the exam to continue her education or to secure a well-paid blue-collar job. From the figures in Panel A, both measures of performance indicate that students of both castes perform much better in English schools, with close to 75% of the English sample and less than half of the Kannada sample passing the exam. Females of both castes are under-represented in the sample of test-takers, suggesting that there are gender-specific selection pressures in secondary school participation. Panel B also makes obvious the different selective pressures on SC/ST and non-SC/ST students in terms of their choice of medium of instruction. While SC/ST students comprise roughly 14% of all English test-takers, they constitute a share of 25% in Kannada. Although returns in the labor market that are both caste- and medium-specific could account for such selection, inter-caste income inequality might also explain the differences in participation. If their parents did not progress past 10^{th} grade or hold a white collar job, SC/ST children would not have experienced much exposure to English and therefore might be reluctant to pursue their education in a "foreign" language. Moreover, the Karnataka government has issued a ban (upheld by the Supreme Court in 1994) on English instruction before the 5^{th} grade in government-funded schools. Thus, any student who has studied English prior to 5^{th} grade must have attended a private primary school that charges tuition (unlike the free government primary schools). The simple model of the previous section provides a test for an alternative explanation for these differences in selection: the better English schools systematically deny admission to SC/ST students.

Panel B also presents the average number of students per school (i.e. the average size of groups in the context of school fixed effects) and provides a breakdown of the share of government, aided, and unaided schools in each medium. The basic regressions outlined in Section 3.7 will also be run separately for each management type to account for heterogeneity in schools' ability and desire to screen students.

4.1.2 School-attendance areas

Because no administrative boundaries restrict individuals from attending a particular school, a definitive measure of a school-attendance area does not exist. In order to test the model's predictions about within-school gaps in the middle of an area's overall performance distribution and their sensitivity to the number of schools across which individuals can sort and to the share of SC/ST individuals in an area, the boundaries of attendance areas had to be defined spatially using information on the location of schools, under the assumption that transportation costs limit a person's choice of school. In separate files, KSEEB and the Department of Public Instruction (DPI) provided village information for all schools sending candidates to the exam in the 2002-2004 period. A private company in Bangalore (Spatial Data Private Limited) created a geo-coded map of Karnataka's Census villages, mapped 94.22% of the KSEEB schools, and attached data from the 2001 Census of India on village attributes such as total population, SC/ST population, and total area to village boundaries. Appendix B describes the procedure used to map additional schools, which brought the percentage of schools in the 2002-2004 period that could be assigned a village up to 98.29. Assigning village locations to schools outside of this period was not so straightforward because of the difficulty of tracking schools across years. Three rounds of redistricting that occurred in Karnataka in 1998, 2000, and 2001 changed the assignment of school identification numbers, and KSEEB could only provide a partial key that matched school identification numbers before and after the rounds of redistricting. Appendix B explains how schools prior to 2002 were assigned village locations and shows why the part of the empirical analysis that requires knowledge of a school's location only employs data from the 2000-2004 period.[31]

Since KSEEB does not supply address information of the test-takers and since schools are assigned to villages and not (x,y)-coordinate locations, it is not possible to create catchment areas based on the exact location of individuals or schools. Instead, a school's village has to serve as the geo-

[31] Knowledge of a school's location is needed when defining quintiles of a school attendance area's school-level performance distribution or when testing the sensitivity of the caste-gap in performance to an area's characteristics such as N or λ.

graphic center of a school-attendance area, and all individuals attending schools in the same village are assumed to come from the same catchment area. In order to demarcate the boundaries of the school-attendance area associated with each village, circles with 3,5,7, and 10km radii were drawn around the center points (centroids) of all villages that contain at least one school with GIS software (ArcMap). The 5km-radii will define the school-attendance areas in the main empirical analysis, but all results presented are robust to the alternative distances.[32] Figure 5 depicts this procedure for a 5km radius in Makodu village in the district of Mysore. All of the darkly outlined villages (and the schools within them) that fall partially within the circle are considered part of the school attendance area associated with Makodu village (and the school within it) which sits in the center of the circle. The school in Makodu is assigned a quintile using all of the schools of the same medium found in its attendance area to form the overall school-performance distribution, and its attendance area assumes the attributes (N, λ) of all of its component villages. Similarly, a school in a village adjacent to Makodu will be in the center of its own school-attendance area and a slightly different set of schools will be used to determine its performance quintile.

Table 4 describes the attributes of school-attendance areas defined by 5km radii from village centroids in the 2000-2004 period. Panel A indicates that though Kannada schools far outnumber schools that teach in English, school attendance areas that contain English schools have much higher concentrations of schools, test-takers, and people than areas that host Kannada schools. This results from the fact that English schools are concentrated in urban environments whereas Kannada schools tend to be more dispersed throughout the state. These figures accord with the maps in Figures 6 - 8 that display the spatial distribution of English and Kannada schools for the areas around the cities of Mysore, Bangalore, and Hubli-Dharwad. Note that the English schools displayed in black mostly appear in clusters, while the gray Kannada schools outside of the cities appear relatively isolated. Thus, the concern about an insufficient number of schools to distinguish between non-discriminatory and discriminatory allocations applies more to the Kannada sample. Panel B provides identical statistics for school-attendance areas that only contain villages with a population less than 1,000,000. This amounts to removing all (50) school attendance areas that intersect the Bangalore Metropolitan "village", which contain 36.39% of all English schools and 5.82% of Kannada schools in the state. This village displays an unusual clustering of both people (2,660,088 in 2001) and schools (734 in 2004), so

[32]Given that most students walk to school or take an auto-rickshaw or local bus, the 3km- and 5km-based definitions of a catchment area are probably more realistic than the ones that use 7km or 10km radii.

to ensure that the sorting processes occurring in this particular area are not driving the main results, each specification of the empirical predictions will also use a sample that excludes these metropolitan school-attendance areas.

4.2 Results

This section first tests for evidence of taste-based discrimination with versions of Equations (1) and (2) adjusted for the fact that KSEEB administers a different exam every year and that the model only makes predictions for the relative performance of SC/ST and non-SC/ST students within a school in a given year (or admission cycle). Given that a student's choice of medium occurred long before her choice of a specific school and that the equilibrium results of the model should hold *within* mediums, all empirical specifications are run separately for English and Kannada schools. Finally, the unequal shares of male and female students taking the exam in both mediums suggest that selection pressures might differ across gender as well (i.e. all females, regardless of their caste, may be drawn from a separate distribution of ability), so the basic empirical model also first includes a gender-caste interaction to test whether the empirical predictions hold within genders. Table 5 provides estimates of the average caste-gap from the following ordinary least-squares and school-year fixed effects regressions,

$$g_{it} = \alpha_0 + \alpha_1 SCST_i + \alpha_2 Female_i + \alpha_3 SC/ST_i * Female_i + \delta_t + \epsilon_{it} \tag{6}$$

$$g_{ijt} = \beta_0 + \beta_1 SCST_i + \beta_2 Female_i + \beta_3 SC/ST_i * Female_i + \delta_{jt} + u_{ijt}, \tag{7}$$

where test-performance g is measured as both the overall pass-status and the total percentage of points earned on the exam. Although females of both castes outperform their males counterparts in the population, the OLS results without school fixed effects (Columns 1,3,5, and 7) confirm that both male and female SC/ST students perform worse than their non-SC/ST counterparts, a finding that is consistent with both the non-discriminatory and discriminatory allocations. In English schools, SC/ST males are close to a 16 percentage point disadvantage in terms of passing (or around 0.37 of a standard deviation), and SC/ST females fall behind non-SC/ST females by a similar magnitude. In terms of the total percentage points earned on the exam, SC/ST males earn an average of 9.4 percent fewer points on the exam (or 0.46 of a standard deviation), with SC/ST females similarly trailing non-SC/ST females. In Kannada schools, SC/ST males score 5.6% fewer percentage points (or 0.33 of a standard deviation less) than their non-SC/ST counterparts, and in the population, SC/ST females deviate from non-SC/ST females by a larger magnitude. Although in terms of the percentage of total points the difference between SC/ST and non-SC/ST males in Kannada schools is smaller than in English, because these students are scoring near the threshold for passing (0.35), the magnitude of the difference in terms of passing rates is still sizeable: SC/ST males display a 12 percentage point disadvantage (or 0.25 of a standard deviation) and the difference between SC/ST and non-SC/ST females is even larger.

The fixed effects results (Columns 2, 4, 6, and 8) provide estimates of the average caste-gap *within* schools and establish the basic premise of the model, that schools care about the test-score potential of their students.[33] Although the SC/ST disadvantage does not disappear entirely as predicted by perfect screening in a non-discriminatory allocation, it does not turn into an SC/ST advantage as predicted by the model that includes discrimination in admissions. In terms of total points earned on the exam, SC/ST males on average score only 2.6 percentage points less than the non-SC/ST males in their school (or 0.13 of a standard deviation), and SC/ST females fall short of their non-SC/ST classmates only slightly more than this. The Kannada results also provide no evidence of discriminatory admissions. SC/ST males still score 3.1 percentage points less than their non-SC/ST counterparts (or 0.18 of a standard deviation), with the difference between SC/ST and non-SC/ST females within schools roughly 1.5 percentage points larger. Taken together, these basic OLS and FE results suggest that schools either directly set lower admissions standards for SC/ST students or, if schools can only imperfectly screen students by their test-score potential, they "statistically discriminate" *in favor* of SC/STs by ignoring the possible disadvantages that accrue while they are in high school.

Since the results for males and females appear qualitatively similar (i.e. neither gender displays evidence of discrimination), the next set of regressions pools genders and shows that these average results hold across periods and remain robust to the inclusion of repeaters and the exclusion of metropolitan areas. Because subsequent regressions that require information on the location of schools (i.e. those that look within quintiles and compare school-attendance areas with different levels of N and λ) can only use data from the 2000-2004 sample, Table 6 demonstrates for both *pass* and *percent* that the basic patterns observed in Table 5 do not change across different period-wise sub-

[33]This smaller caste-gap estimated within schools mirrors results found in the school-choice literature that suggest that students will sort by academic ability when given the chance (Epple and Romano (2002), Nechyba (2003), Cullen, Jacob, and Levitt (2005), Hsieh and Urquiola (2006))

samples of the data - specifically that the 2000-2004 period does not represent a departure in terms of the sorting processes that allocate students to schools. In English schools, although SC/ST students are gaining relative to non-SC/ST students in the population in periods 2 (*pass* and *percent*) and 3 (*pass*),[34] the average within-school gap remains constant. Columns 5 - 8 repeat this exercise for Kannada schools. In terms of pass rates and the percentage of total points earned on the exam, SC/ST students in the population lose ground relative to non-SC/ST students in period 3, but this translates into a only a minor increase of 0.7 of a percentage point in pass rates in the within-school gap for that period and does not affect within-school differences in terms of the total points score on the exam.

Table 7 presents results for *percent* for samples that include repeaters (Columns 1 -4) and exclude metropolitan school-attendance areas (Columns 5 - 8) in the 2000-2004 period. The basic patterns do not change when within-school comparisons are made even with individuals admitted in different years,[35] and the Bangalore metropolitan area alone is not driving the main results.

The previous regressions pooled together all schools in an area's performance distribution. To address the possibility that the average caste-gap across all schools conflates SC/ST performance advantages in the middle of the distribution with disadvantages in the tails that arise from the lower supports of $F^L(z)$ relative to $F^H(z)$, the regressions of Table 8 take into account the quintile of school j in the overall performance distribution of its school attendance area and allows the caste-gap to vary by quintile. Columns 1 - 4 present estimates from areas in the the 2000-2004 sample that contain mutually exclusive quintiles and thus exclude all areas with less than 5 schools. In fact, in this sample, the average number of English schools in areas that contain at least one English school is 77, while the corresponding number for Kannada schools is 16. The fixed effects estimates indicate that SC/ST students do not outperform their non-SC/ST classmates in any quintile of schools' average performance distribution. Although the within-school caste-gaps are smaller in the 1^{st} and 2^{nd} quintiles in English schools relative to the 5^{th}, SC/ST students still do not hold an advantage. Again, to ensure that the impressive concentrations of schools in metropolitan areas are not driving the results, Columns 5- 8 replicate the results using a sample that excludes these areas, in which the average numbers of English and Kannada schools per area drop to 16 and 10, respectively. As a group, SC/ST students never demonstrate better performance than their non-SC/ST classmates in

[34]If this trends results from a higher rate of switching mediums for non-SC/ST students because of increasing returns to English education, then it might make more sense to state "although non-SC/ST students are losing ground relative to SC/ST students...".

[35]To the extent that a school can deny admission in the subsequent year to its students that failed the exam, however, repeaters can be considered to be a part of the same admissions cycle as the first-time test-takers.

any quintile.

Table 9 repeats this exercise separately for government, aided, and unaided schools to account for heterogeneity in schools' ability and desire to screen students. Given that government schools cannot administer admissions tests or conduct interviews, it might be difficult to detect discrimination in a government school if it does not discriminate by setting different admissions thresholds but rather arbitrarily denies admission to some SC/ST students, irrespective of their test-score potential. To the extent that government schools rank below the private schools in terms of quality (see the "Observations" row in Tables 10 and 11), however, discrimination practiced by the other types of schools should manifest itself in government schools' upper thresholds since all of the SC/ST students shut out of the better schools would have to attend the lower ranked government schools. The within-school caste-gaps estimated in Columns 2, 4, 6, 8, 10, and 12 provide no evidence of taste-based discrimination in any type of school. Moreover, the gaps do not appear to wildly differ across types. Although not reported here, estimates from the non-metro sample do not differ much.

Tables 10 and 11 presents results from examining differences between percentiles of the SC/ST and non-SC/ST distributions within schools. This exercise requires less stringent assumptions on the shapes of $F^H(z)$ and $F^L(z)$ but at the same time requires that schools contain at least 10 SC/ST students, which dramatically reduces the number of schools in the sample. Panel A in both tables reports results from all management types combined. In all percentiles, SC/ST students still perform slightly worse than their non-SC/ST classmates. In both mediums of instructions and all along an area's school performance distribution (i.e. in the top, middle, and bottom schools), the difference in lower thresholds (i.e. differences at the 10^{th} percentile) is always smaller than the corresponding difference in the upper thresholds, suggesting that schools do set roughly similar admissions thresholds for SC/ST and non-SC/ST students but that non-SC/ST students comprise a larger share of the highest ability students that schools admit. Panels B, C, and D confine their attention to government, aided, and unaided schools, respectively and show that qualitatively similar results hold within management types, although the small samples size of unaided English schools at the bottom of areas' performance distributions makes it difficult to estimate the difference in thresholds in these schools with any precision.

Figures 9 and 10 depict the results from the rescaled quantile regressions of Equation (4). Aside from swings in the 5^{th} and 95^{th} quantiles, the quantile estimates of Figure 8 suggest that OLS regressions (or a focus on differences in means within schools) provide a good approximation for average differences in various quantiles. Moreover, the grey-shaded confidence intervals of the quantile estimates lie well below zero. Figure 9 distinguishes among top, middle, and bottom schools (with schools in their area's fifth quintile classified as *top*, schools in the first quintile as *bottom*, and schools in the other quintiles as *middle*) and shows that qualitatively similar results hold along areas' performance distributions.

The regressions of Table 12 show how the differences in percentiles within schools respond to changes in the opportunity to sort and changes in a school-attendance area's SC/ST share, where an area's population density proxies for N in the model and an area's medium-specific share of SC/ST test-takers represents λ. Results would not change considerably if the number of schools were substituted for population density or if the overall SC/ST share of the entire population replaced the test-taker's share. Given that most English schools operate in areas that offer a considerable amount of school-choice, it is not surprising to see that increases in N are generally not associated with a narrowing of the caste-gap within schools. Similarly, increases in a school-attendance area's SC/ST share area are associated with larger (i.e. more negative) within school differences between SC/ST and non-SC/ST students, contrary to what the model predicts for how discriminatory schools would admit students when the SC/ST share increases. In Kannada schools, although marginal increases in N reduce the caste-gap within schools, the maximum implied increase from moving from an area with a population density equal to the 10^{th} percentile of population density (116 people per square kilometer) to an area with the density of the 90^{th} percentile (662 people per square kilometer) is 0.0001 percentage points (for the top percentile). Likewise, marginal increases in an area's SC/ST share are associated with smaller (less negative) caste-gaps, but the increase in SC/ST share that corresponds to the change involved in moving from the 10^{th} to the 90^{th} percentile of SC/ST test-taking share (from 9.6 to 40.2) results in a narrowing of at most 0.0063 in the top percentile. These results suggest that insufficient numbers of schools and low SC/ST shares are not obscuring evidence of discrimination. This paper, however, is silent on the determinants or correlates of discrimination. There might be all sorts of requisite changes in attitudes towards SC/STs (and therefore changes in the relative values of $\frac{\partial^2 v}{\partial z^2}$ and $\frac{\partial^2 v}{\partial \lambda^2}$ - the tradeoff between average test scores and SC/ST shares in a school) that might change with population density and the area's SC/ST share. If the tendency of schools to discriminate were negatively correlated with N ("not in the cities") or the area SC/ST share λ ("more power"), then caste discrimination would only be present precisely in the areas where the non-discriminatory and discriminatory allocations cannot be distinguished (areas with low school counts or areas with low SC/ST shares in the population).

4.3 Discussion

All of these results - the OLS specifications, the adjustments for a school's position in the performance distribution, the quantile regressions - indicate that SC/STs do not face discrimination in the admissions process. The specific pattern - namely, the smaller caste-gaps at the bottom of the distribution within schools (the 10^{th} percentile) relative to the top (the 90^{th} percentile) - is consistent with schools' setting a lower bound of test-score potential (an admissions threshold) and individuals' displaying different distributions of test-score potential across castes - namely, SC/STs' displaying a lower right-hand support.[36]

The interpretation of the non-positive coefficient on caste as evidence against admissions discrimination, however, relies on certain assumptions made in the model on the objective functions of schools and individuals and on the test-score production process. How plausible are these assumptions? First, the model (and the inference made in the empirical analysis) assumes that schools always want to maximize test scores. If schools randomly discriminated against SC/STs and therefore denied them admission without regard to their test-score potential, then a test that measures the degree of selection exhibited by SC/STs within schools would fail to find discrimination in admissions even if it were occurring. The fact that schools are publicly ranked in the newspaper *should* provide strong incentives for schools to admit students with the highest test-score potential. Although given data limitations it would be impossible to say with certainty that students are sorted across schools according to their incoming test-score potential, these data can be used to assess the degree of sorting with respect to ex-post potential - that is, whether or not students appear sorted according to their test scores. Table 13 offers some evidence on the degree of ex-post sorting within school-attendance areas. Within each area, schools are assigned separate ranks for their 25^{th}, 50^{th}, and 75^{th} percentiles (the actual rank when the area contains ten schools or less, the quintile of the area's entire performance distribution when there are more than 10 schools). Schools are then assigned an indicator equal to 1 if their three ranks are equivalent. Thus, in an area with five schools, if a school's ranks for the 25^{th}, 50^{th}, and 75^{th} percentiles are (4,4,4) it receives a value of 1, but if its ranks are (4,4,5) or (1,

[36]If, of course, attributes such as demeanor play a large role in admissions decisions in addition to any scores on an entrance exam, it would be difficult to imagine how schools could actually pinpoint an exact lower bound of incoming test-score potential.

3, 5) then it receives a value of 0.[37] These indicators are then averaged within the school-attendance area to get a measure of the fraction of schools in an area that have not changed ranks. The results of Table 13 appear to indicate that there is a fair amount of sorting with respect to realized outcomes. Although, for example, the typical school-attendance area with 10 Kannada schools or less only has 52% of its Kannada schools that preserve their exact rank across the 25^{th}, 50^{th}, and 75^{th} percentiles, this lower percentage appears to stem from slight deviations from pure sorting (e.g. schools with ranks such as (2, 2, 3) or (3, 4, 3) in areas with five schools) rather than from random sorting, as the higher percentages suggest when only the 25^{th} and the 50^{th} or the 50^{th} and the 75^{th} percentiles are compared.

Another critical assumption of the model is individuals' desire to seek schools offering the highest level of school inputs. If some students sought schools of lower quality with the intention of purchasing extra-curricular inputs (e.g. after-school tutoring) in order to maximize their test scores and if non-SC/STs are more likely to choose schools in this fashion, then the discrimination test proposed in this paper might also fail to find evidence of admissions prejudice even when it is occurring. In this case, it would be the non-SC/STs that would always appear as a select sample within schools even if schools paid no attention to caste. Selection pressures on SC/STs stemming from discrimination in admissions would then have to overcome this initial selection of non-SC/STs. The data used in this paper cannot comment on this issue, but the long waiting lists that schools maintain suggest that individuals are indeed seeking schools of the highest quality.

The third assumption maintained throughout the analysis has been the monotonicity of the test-score production process with respect to incoming test-score potential. Failure of this assumption, however, would only threaten a result of no discrimination if very a very specific kind of cross-caste peer spillovers were occurring - namely, SC/STs at the top and bottom of the distributions within schools only pulled up non-SC/STs to their level and had no effect on students of their own caste (or non-SC/STs only brought down SC/STs to their level).

Finally, Table 14 addresses the issue of differential retention and drop-out. In the past, school managements administered their own "selection exams" to determine whom they would send to the state exam, with the weakest students typically retained if they demonstrated a high likelihood of failing (Karanth, 2006). Thus, when these students finally took the exam, they were classified as first-

[37] As a result, this is a rather stringent test since it penalizes slight deviations from pure ability sorting (4,4,5) in the same way that it penalizes what appears to be totally random sorting (1,3,5).

time test-takers, despite the fact that they had repeated the 10^{th} grade. If this were still occurring, this would only pose a problem (i.e. this would only mask evidence of discrimination) if non-SC/ST students were retained more - that is, if the weakest non-SC/ST students were prevented from taking the exam by their schools and the non-SC/ST threshold was therefore biased towards the higher threshold that applied to SC/ST students. To the extent that SC/ST and non-SC/ST students enter their high school at the same age, age differences across castes at the time of the exam would give an indication of the extent of differential retention. If non-SC/ST students were retained more, they should appear older than the SC/ST students in their exam-cohort. Table 3, however, showed that it is the non-SC/ST students in the sample who are younger, and Table 13 shows that this relationship holds within schools as well. This evidence on the age gap in fact further counters the possibility that schools are discriminating against SC/STs by holding them to higher admissions standards. The 1991 Census data suggests that SC/ST students in Karnataka begin their formal schooling later than non-SC/ST students (the participation gap between SC/ST and non-SC/ST children declines from ages 5 to 10). To the extent that late-starters would demonstrate lower levels of test-score potential (or test-taking ability) at the time of high school admission, this age gap in primary school does not guarantee that within schools SC/ST students would always be older than their non-SC/ST classmates. In fact, if age at entry and test-score potential (or test-taking ability) were sufficiently negatively correlated, then in a discriminatory allocation, even without differential retention, SC/ST students would appear younger within schools. Differential retention then would only exacerbate this difference, making SC/ST students look even younger than the non-SC/ST students in their school who take the exam with them.

If the weakest students were instead expelled from the school and forced to drop out, this would again bias the non-SC/ST threshold towards the SC/ST threshold. These types of censoring associated with retention and drop-out, however, would not alter differences in the upper thresholds that define schools. Because of the recursive nature of admissions, the top students in the schools in the common support of the two ability distributions would be SC/STs in the discriminatory equilibrium. The lack of a positive $SCST_i$ coefficient even at the 90^{th} percentile suggests that differential retention or drop-out are not obscuring evidence of discrimination in Karnataka's schools.

5 Conclusion

This paper shows that the concentration of SC/ST students in the worst high schools in Karnataka cannot be explained by discriminatory admissions policies on the part of schools. Data on test-scores from a sample of 5.7 million individuals who took the state-wide matriculation exam from 1996-2004 and geographic information on the location of schools permits a comparison of the SC/ST and non-SC/ST distributions of scores within schools along the distribution of school-quality and shows that SC/ST and non-SC/ST students do not face different admissions thresholds with respect to test-score potential demonstrated prior to high school. Specifically, the results are consistent with a model of school admissions in which schools only care about maximizing their average test-score and do not care about the caste-composition of their schools. It is consistent with a model of taste-based discrimination against SC/STs in admissions only if schools simultaneously overestimate their test-score potential and fail to recognize possible differences in gains made during high school that might arise from discrimination in the classroom, differential access to complementary inputs, or differences in motivation.

It would be difficult to generalize these results to the rest of India or to the full spectrum of castes without knowing how Karnataka ranks relative to other states in terms of tolerance towards SC/STs or if discrimination amongst adjacent castes in the caste hierarchy is stronger than discrimination towards SC/STs. These results, however, are still remarkable given that the SC/ST status continues to limit the scope of other social interactions in Karnataka such as marriage and the use of public water sources. They are also instructive for policies that aim to reduce inter-caste inequality in Karnataka between SC/STs and the rest of the population. First, regardless of whether or not schools discriminate, the current allocation of students to schools suggests that affirmative action policies that apply only after high school may be an ineffective tool for reducing inequality if only very few SC/ST individuals - the so-called "creamy layer" in the language of the Indian affirmative action debate - can attend the good schools that will prepare them adequately for the exam and allow them to take advantage of quotas in universities and in well-paying government service jobs. The pass rate, for example, in schools in the bottom two quintiles in 2004 that serve over 56% of the SC/ST population is less than 47%. Second, to the extent that the distribution of test-taking ability demonstrated prior to 8^{th} grade does not result from discriminatory admissions in primary school and to the extent that secondary schools do not need to take discrimination within the classroom into account when making their admissions decisions, the absence of prejudiced admissions practices suggests that a policy aiming to increase SC/ST access to school quality need not target caste and expand affirmative action to secondary schools. Given recent reactions in India to reservation-hikes in public universities, it is easy to conjecture that such expansions would generate a considerable amount of social conflict. Investments in improving school quality in the worst primary and secondary schools, however, would disproportionately benefit SC/ST students without explicitly targeting them, as would means-tested transfers designed to equalize access to complementary inputs such as after-school tutoring.

It would also be unwise to use this paper's results to inform affirmative action policies for primary school and universities - institutions that might face different incentives to maximize the academic performance of their students. There is no state-wide exam that students are required to take before the matriculation exam in 10^{th} grade, and universities administer their own exams before letting their students graduate. It might be the case that secondary schools are indeed prejudiced against SC/STs but performance incentives created by the ranking of schools according to a common test mitigates any incentives to discriminate. Thus, the finding that secondary schools in Karnataka do not discriminate against SC/STs in admissions in the absence of an applicable affirmative action policy may only tell us that schools do not exhibit taste-based discrimination when they are evaluated according to explicit, publicly observable performance criteria.

References

[1] Alesina, Alberto, Reza Baqir, and Caroline Hoxby (2004), Political Jurisdictions in Heterogenous Communities. *Journal of Political Economy*. 112 (2), pp. 348-96.

[2] Akerlof, George (1976), The Economics of Caste and of the Rat Race and Other Woeful Tales. *Quarterly Journal of Economics*. 90 (4), pp. 599-617.

[3] Akerlof, George (1997), Social Distance and Social Decisions. *Econometrica*. 65 (5), pp. 1005-1027.

[4] Anwar, Shamena and Hanming Fang (2006), An alternative test of racial prejudice in motor vehicle searches: theory and evidence, *American Economic Review* 96 (1), pp. 127 - 151.

[5] Arrow, Kenneth. (1972), Models of Job Discrimination. *Racial Discrimination in Economic Life*.(Anthony H. Pascal, ed.), RAND Corporation, pp. 82-102.

[6] Becker, Gary (1957), *The Economics of Discrimination*. University of Chicago Press.

[7] Becker, Gary (1993), The evidence against banks doesn't prove bias *Business Week*. April 19, 1993.

[8] Becker, Gary and Kevin M. Murphy (2000), *Social Economics: Market Behavior in a Social Environment*. Harvard University Press, pp. 47-73.

[9] Behal, Suchitra (2002) Caste cruelty makes school a nightmare. *The Times Educational Supplement*, April 26

[10] Berry Cullen, Julie, Brian A. Jacob, and Steven D. Levitt (2005), The impact of school choice on student outcomes: an analysis of the Chicago Public Schools. *Journal of Public Economics*. 89, pp. 729-760.

[11] Boustan, Leah Platt (2006), Inside the black box of "white flight": the role of suburban political automony and public goods.

[12] Cooley, Jane (2006), Desegregation and the achievement gap. Do diverse peers help?

[13] Epple, Dennis, David Figlio, and Richard Romano (2004), Competition between private and public schools: testing stratification and pricing predictions. *Journal of Public Economics*. 88, pp. 1,215-45.

[14] Department of Public Instruction, Government of Karnataka (2003), Citizens Charter.

[15] Department of Public Instruction, Government of Karnataka (2004), Application to open a private school.

[16] Government of Karnataka (2004), Note on abolition of capitation fee in private institutions and taking action against such institutions.

[17] Government of Karnataka (1994), The Karnataka Scheduled Castes, Scheduled Tribes and Other Backward Classes (Reservation of Seats in Educational Institutions and of Appointment or Posts in the Services under the State) Act of 1994.

[18] Hsieh, Chang-Tai and Miguel Urquiola (2003), When schools compete, how do they compete? An assessment of Chile's nationwide school voucher program. NBER Working Paper 10008.

[19] Jaffrelot, Christophe (2006), The impact of affirmative action in India: more political than socioeconomic. *India Review*. 5(1), pp. 173-189.

[20] Jenkins, Laura Dudley (2003), *Identity and Identifcation in India: Defining the Disadvantaged*. RoutledgeCurzon.

[21] Karabel, Jerome (2005), *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton*. Houghton Mifflin Company, New York.

[22] Karanth, P.S. (2006), Interview.

[23] Knowles, John, Nicola Persico, and Petra Todd (2001), Racial bias in motor vehicle searches: theory and evidence. *Journal of Political Economy*. 109(1), pp. 203-229.

[24] Koenker, Roger and Gilbert Basset, Jr. (1978), Regression Quantiles. *Econometrica*. 46(1), pp. 33-50.

[25] Lundberg, Shelly and Richard Startz (1983), Private discrimination and social intervention in competitive labor markets. *American Economic Review*, 73(3), 340-347.

[26] Munshi, Kaivan and Mark Rosenzweig (2005), Traditional institutions meet the modern world: caste, gender, and schooling choice in a globalizing economy. *American Economic Review*, forthcoming.

[27] Nechyba, Thomas (2003), Introducing school choice into multi-district public school systems. *The Economics of School Choice* (Caroline Hoxby, ed.) University of Chicago Press, pp. 145-94.

[28] Pascal, Anthony H. and Leonard A. Rapping (1972), The economics of racial discrimination in organized baseball. *Racial Discrimination in Economic Life*.(Anthony H. Pascal, ed.), RAND Corporation, pp. 119-156.

[29] Ramachandran, Vimala (2002), Beyond the numbers. *The Hindu*, February 24.

[30] Sainath, P.(1999), This is the way they go to school. *The Hindu*, November 28.

[31] Sethi, Rajiv and Rohini Somanathan (2004), Inequality and segregation. *Journal of Political Economy* 112(6), pp. 1296-1321.

[32] Srinivas, M. N.(1954), Varna and caste. Reproduced in *Collected Essays*, 2002, pp.166-172.

[33] Srinivas, M. N.(1968), Mobility in the caste system. Reproduced in *Collected Essays*, 2002, pp.187-199.

[34] Tilak, Jandhyala (2002), Determinants of household expenditure on education in rural India. National Center for Applied Economic Research (New Delhi), Working Paper Series 88.

6 Appendix A: Excerpts from the 2005 English, Math, and Social Science exams

36. Who rang the doorbell when Rodney Hobson was in the kitchen ?
 (A) A friend (B) A salesman
 (C) His wife (D) A neighbour.

37. How does Mr. Mumford describe the veal ?
 (A) A half-inflated rugby football (B) Considerably inflammable
 (C) The very devil (D) A passable piece of beef.

38. Who invaded England in 793 A. D. after the West-Germanic tribes settled down ?
 (A) The Normans (B) The Scandinavians
 (C) The Angles (D) The Saxons.

39. A Genito-urinary surgeon operates upon the
 (A) heart and lungs (B) kidneys, bladder or ureter
 (C) brain, spinal cord and nerves (D) ears, nose and throat.

40. Why is a patient often in worse shape after the operation than before ?
 (A) Due to prolonged anaesthesia and extensive surgical trauma
 (B) Due to bed-confinement
 (C) Due to too much of acid in the blood
 (D) Due to a drop in blood pressure.

41. What does the poem, "The Character of a Happy Life" depict ?
 (A) That man is happy who likes a pious but grand life
 (B) The quiet philosophy of a man who knows how to live
 (C) Man's mind is troubled by greed and envy.
 (D) Man is touched by praise and blame.

61. How was Radhakrishnan's send-off from Mysore unique ?

62. "By eight O'clock the fire had burned low." Explain.

63. What does Cowper mean by, 'I reflect on the frailty of man and his joys' ?

64. Where does Liberty live, according to Lord Byron ?

25. $\Delta ABC \; ||| \; \Delta DEF$. The areas of these triangles are 9 sq.cm and 25 sq.cm respectively. If AB = 15 cm, then the corresponding side DE =

(A) 9 cm (B) 5 cm

(C) 25 cm (D) 15 cm.

26. The biggest chord of a circle is

(A) radius (B) diameter

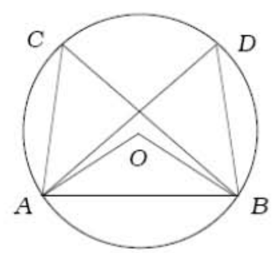
(C) semi-arc (D) circumference.

27. Circumference and diagonal of a circle are c and d. Then the value of $\frac{c}{d}$ is

(A) π (B) radius

(C) semi-circumference (D) 2π.

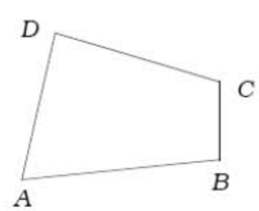
28. O is the centre of a circle. $\angle AOB = 100^\circ$. Then the values of $\angle ACB$ and $\angle ADB$ are :



(A) 50° and 60° (B) 50° and 50°

(C) 60° and 60° (D) 60° and 50°.

29. In the cyclic quadrilateral $ABCD$, $\angle ABC = 95^\circ$, $\angle DAB = 70^\circ$. Then the values of $\angle ADC$ and $\angle BCD$ are



(A) 70° and 95° (B) 95° and 70°

(C) 85° and 110° (D) 110° and 85°.

17. The special power that is enjoyed by the permanent members of Security Council is

(A) co-ordinating the Secretariat with others

(B) supervising power during emergency

(C) power to judge international affairs

(D) veto power.

18. The hill tribal refugees that are pouring into India from Chittagong are known as

(A) Singhalese (B) Chakmas

(C) Nepalis (D) Tibetans.

19. Biological Weapon Convention came into effect in the year

(A) 1975 (B) 1985

(C) 1995 (D) 1965.

20. London is the headquarters of Commonwealth of Nations whereas the headquaters of SAARC is at

(A) Nepal (B) Maldives

(C) Bhutan (D) Sri Lanka.

21. The highest recorded rainfall is at

(A) Mawsynram

(B) Cherrapunji

(C) The Western Ghats

(D) The Northern region of Himalayas.

22. Gir forest of Saurashtra is the natural habitat of the lions whereas the Sunderbans are the habitat of

(A) Leopards (B) Tigers

(C) Elephants (D) Bears.

23. The National Forest Policy was introduced in the year

(A) 1952 (B) 1962

(C) 1982 (D) 1972.

24. Which one of the following is the longest dam of the world ?

(A) Tungabhadra (B) Kosi

(C) Hirakud (D) Chambal.

25. To cultivate paddy 25°C temperature is required, whereas to cultivate wheat the temperature required at sowing time is

(A) 10°C – 15°C (B) 15°C – 20°C

(C) 20°C – 25°C (D) 30°C – 35°C.

•

Note: Students are expected to know any quoted passages or references to stories or poems. They are not given any material as context.

7 Appendix B

7.1 The location of schools

The mapping between the villages listed in the KSEEB/DPI address databases and Census villages was not always perfect. Much of KSEEB/DPI village information included the names of *revenue* villages, pin-codes (or postal codes), major landmarks near the school (e.g. Gersoppa Falls or Ranganathaswamy Temple), and village spellings that departed from those found in the list of villages provided by the Census (most likely because of the numerous ways the Kannada script can be transliterated into English). The spatial data company was able to place 94.22% of the schools in villages on the map. Further mapping had to be done on a school-by-school basis by (i) using lists of revenue villages and their corresponding Census villages on the website of the Rural Development and Panchayat Raj Department of the Government of Karnataka, (ii) looking up pin-codes on the Central Excise and Service Tax website (Government of India), and (iii) Google searches of village landmarks. When either (i) or (ii) yielded multiple villages (i.e. when a revenue village or pin-code corresponded to more than one Census village), the school in question was placed in the most population-dense village if all the Census villages were contiguous. Otherwise, its village was treated as missing. This brought the percentage of 2002-2004 schools that could be mapped up to 98.29.

7.2 Matching schools across years

Schools are uniquely identified in the KSEEB data by their medium of instruction and a school code.[38] The code contains a letter component that identifies its district, followed by a numerical component that identifies the particular school within the district. In 1998, 2000, and 2001, Karnataka implemented three rounds of redistricting that ultimately split its 20 districts into 32. Unfortunately, the Board also changed both the letter and number components of its codes in the affected districts after each round of redistricting. Take the district of Mysore, for example. From 1996 to 1998, Hardwick High School was assigned the code E050 and Sri Mahadeshwara High School the code E076. When Mysore split into the separate districts of Mysore and Chamarajanar, the code for Hardwick in Mysore became EE038 and the code for Mahadeshwara in Chamarajanagar became EA019. The re-

[38]A single management can run two separate schools, often in the same building, one offering instruction entirely in English, the other instruction in Kannada. Since the management must employ two different sets of teachers and since the student populations generally do not mix, these schools are treated as separate schools, although they share a school code.

assignment of codes was not altogether straightforward, and a key linking current school codes with those that prevailed before the 1998 and 2001 redistricting was necessary to link schools across years. The key, however, supplies only a *partial* mapping, so some pre-2002 schools cannot be matched to current schools and therefore cannot be assigned a village or a management type since KSEEB also only provided information on management types for schools in the 2002-2004 period.[39] Table A.1 shows the missing rates for various years. A large discontinuity appears in 1999 because most of the redistricting occurred in 1998, and it was therefore difficult to distinguish schools with reassigned school codes from genuinely new schools that appeared in this year. The more schools missing from a school-attendance area, the less reliable will be quintile-assignment for schools that do appear in these areas. Therefore, the part of the empirical analysis that distinguishes schools by their placement in their school-attendance area's performance distribution only employs data from the 2000-2004 period.

[39]The assistant of the then director of the Board who provided the key claims that even the Board does not have the missing data since it was lost while switching their database from Foxprow to Oracle.

Table A.1: Schools missing village locations and management types

Year	Percentage of schools missing village location	Percentage of schools missing management type
1996	27.97	26.68
1997	29.04	27.75
1998	29.41	28.09
1999	15.15	13.65
2000	5.27	3.9
2001	3.3	2.04
2002	1.3	0.02
2003	1.26	0
2004	1.29	0.01

Table 1: Inter-caste inequality in Karnataka

	SC/ST	Non-SC/ST
Education		
Completion rate of individuals aged 25-59 years		
Primary	0.263	0.523
Secondary	0.131	0.298
Tertiary	0.040	0.108
Observations	4,610,542	17,012,656
Assets		
Percent of households that:		
Use banking services	0.237	0.452
Own a television	0.207	0.421
Own a two-wheeler	0.057	0.172
Own an automobile	0.010	0.037
Observations	2,460,656	7,771,477

Source: Census of India, 2001

Primary school covers grades 1-4; secondary school, grades 8 - 10; and tertiary school, any schooling after grade 12 (diplomas, bachelor's degress, and post-graduate education).

A two-wheeler is a scooter, motor-bike, or moped.

Table 2: Simulated caste-gaps from two normal distributions

Variable:	Probability that statistic is higher for the H group when L-share is 20%											
Statistic:	Mean				Min				Max			
Quartile:	1	2	3	4	1	2	3	4	1	2	3	4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\mu^H = \mu^L = 0$	0.507	0.501	0.492	0.492	0.202	0.201	0.200	0.202	0.795	0.798	0.801	0.797
$\mu^H = \mu^L + 0.25$	0.764	0.619	0.630	0.798	0.173	0.196	0.228	0.359	0.899	0.830	0.803	0.777
$\mu^H = \mu^L + 0.50$	0.883	0.719	0.744	0.966	0.136	0.193	0.252	0.554	0.956	0.863	0.815	0.758
$\mu^H = \mu^L + 0.75$	0.936	0.795	0.836	0.999	0.106	0.169	0.258	0.731	0.979	0.898	0.827	0.738

The H-distribution consists of 800 draws from a normal distribution with unit variance; the L-distribution, 200 draws from a normal distribution with unit variance.
Quartile 1 refers to the lowest quartile, Quartile 4 the highest.
In the identical distributions case, each had a mean of zero. In all other cases, only the L-distribution has a mean of zero.
Results are from 10,000 replications.

Table 3: Descriptive statistics of individuals and schools

Medium:		English			Kannada		
Sample:		Full sample	SC/ST	Non-SC/ST	Full sample	SC/ST	Non-SC/ST
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A:	**Individuals**						
	Pass	0.748	0.610	0.766	0.465	0.359	0.496
		(0.434)	(0.002)	(0.000)	(0.499)	(0.001)	(0.000)
	Percent	0.565	0.480	0.577	0.386	0.338	0.399
		(0.203)	(0.001)	(0.000)	(0.170)	(0.000)	(0.000)
	Age	15.339	15.462	15.323	15.659	15.800	15.618
		(0.811)	(0.003)	(0.001)	(1.046)	(0.001)	(0.001)
	Female	0.458	0.440	0.461	0.432	0.388	0.444
		(0.498)	(0.002)	(0.001)	(0.495)	(0.001)	(0.000)
	Observations	855,144	99,223	755,921	3,479,215	775,171	2,704,044

Panel B:	**Schools**	**English**	**Kannada**
	Students per school	83.071	92.342
		(0.666)	(0.328)
	SC/ST share	0.142	0.249
		(0.001)	(0.001)
	Government share	0.396	0.513
		(0.006)	(0.003)
	Aided share	0.226	0.305
		(0.005)	(0.003)
	Unaided share	0.363	0.173
		(0.006)	(0.002)
	Observations	11,005	49,998

Columns 1 and 4 of Panel A present standard deviations in parentheses.
Columns 2, 3, 5, and 6 and the columns of Panel B present standard errors in parentheses.
Observations in Panel B are school-years.
Statistics are for the 1996-2004 sample which does not include repeaters, except for the rows that contain shares by management type. These rows only contain observations from 2000-2004.
The variable *percent* represents the total percentage of points (out of 625) earned by a test-taker.
The variable *pass* $= 1$ if *percent* ≥ 35 and the raw score in each subject is at least 30 points.

Table 4: Descriptive statistics of school-attendance areas

Medium:	English	Kannada
Panel A: All areas	(1)	(2)
Number of schools	30.845	12.445
	(1.253)	(0.262)
Number of test-takers	2300.716	888.889
	(95.156)	(13.996)
SC/ST share of test-takers	0.089	0.248
	(0.001)	(0.001)
Population density	2198.721	870.468
(population per square kilometer)	(84.054)	(28.882)
SC/ST population share	0.194	0.220
	(0.001)	(0.001)
Observations	7,637	22,769
Panel B: Excluding areas containing metropolitan villages:		
Number of schools	6.827	7.707
	(0.150)	(0.070)
Number of test-takers	483.782	651.839
	(12.757)	(6.184)
SC/ST share of test-takers	0.086	0.247
	(0.001)	(0.001)
Population density		
(population per square kilometer)	653.564	347.362
	(8.649)	(3.190)
SC/ST population share	0.198	0.221
	(0.001)	(0.001)
Observations	7,277	22,409

All figures are calculated for areas that contain at least one English school (Column 1) or at least one Kannada school (Column 2) in the 2000-2004 period.
The variables *number of school, number of test-takers,* and *SC/ST share of test-takers* refer to medium-specific figures.
Villages are classified as metropolitan if their population exceeds 1,000,000.
A school-attendance area is defined as a circle with a 5km radius around the centroid of a village that contains at least one school.
Observations are area-years in the 2000-2004 period that contain at least one school.

Table 5: The average caste-gap in performance

Medium:	English				Kannada			
Dependent variable:	Pass		Percent		Pass		Percent	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SC/ST	-0.158	-0.039	-0.094	-0.026	-0.124	-0.075	-0.056	-0.031
	(0.007)	(0.002)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Female	0.044	0.042	0.047	0.042	0.045	0.029	0.031	0.021
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
SC/ST*Female	0.007	-0.001	-0.003	-0.006	-0.038	-0.037	-0.014	-0.015
	(0.010)	(0.003)	(0.005)	(0.001)	(0.002)	(0.001)	(0.001)	(0.000)
Constant	0.713	0.733	0.508	0.549	0.396	0.473	0.336	0.385
	(0.009)	(0.001)	(0.005)	(0.000)	(0.004)	0.000	(0.001)	(0.000)
Observations	855,144	855,144	852,836	852,836	3,479,215	3,479,215	3,457,321	3,457,321

Standard errors in parentheses are clustered at the level of the school-year
Columns 2, 4, 6, and 8 include a set of school-year dummies.
Columns 1, 3, 5, and 7 include a set of year dummies.
The constant term in the FE regressions represents the average of all observations' school-year fixed effects.

Table 6: The average caste-gap across periods

Medium:	English				Kannada			
Dependent variable:	Pass		Percent		Pass		Percent	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SC/ST	-0.181	-0.045	-0.103	-0.030	-0.128	-0.086	-0.061	-0.038
	(0.013)	(0.004)	(0.006)	(0.001)	(0.002)	(0.001)	(0.001)	(0.000)
Period 2	0.033	–	0.051	–	0.081	–	0.057	–
	(0.006)		(0.004)		(0.003)		(0.001)	
Period 3	0.030	–	0.047	–	0.158	–	0.080	–
	(0.006)		(0.004)		(0.003)		(0.001)	
SC/ST * Period 2	0.037	0.006	0.011	0.002	-0.008	-0.003	-0.001	0.000
	(0.016)	(0.005)	(0.008)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)
SC/ST * Period 3	0.035	0.007	0.009	0.003	-0.030	-0.007	-0.005	0.000
	(0.015)	(0.005)	(0.007)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)
Constant	0.744	0.753	0.542	0.569	0.413	0.485	0.352	0.394
	(0.005)	(0.000)	(0.003)	(0.000)	(0.002)	(0.000)	(0.001)	(0.000)
Observations	855,144	855,144	852,836	852,836	3,479,215	3,479,215	3,457,321	3,457,321

Standard errors in parentheses are clustered at the level of the school-year
Columns 2, 4, 6, and 8 include a set of school-year dummies.
Period 1: 1996-1998
Period 2: 1999-2001
Period 3: 2002-2004
The constant term in the FE regressions represents the average of all observations' school-year fixed effects.

Table 7: The average caste-gap in the repeat and non-metro samples

Dependent variable:	Percent of total points							
Sample:	Includes repeaters				Excludes metropolitan areas			
Medium:	English		Kannada		English		Kannada	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SC/ST	-0.107	-0.030	-0.058	-0.035	-0.092	-0.025	-0.065	-0.038
	(0.003)	(0.001)	(0.000)	(0.000)	(0.006)	(0.002)	(0.001)	(0.000)
Constant	0.508	0.532	0.341	0.377	0.586	0.588	0.406	0.425
	(0.005)	(0.000)	(0.001)	(0.000)	(0.005)	(0.000)	(0.001)	(0.000)
Observations	1,015,336	1,015,336	4,639,723	4,639,723	194,489	194,489	1,192,822	1,192,822

Standard errors in parentheses are clustered at the school-year level.
Columns 1, 3, 5, and 7 include year dummies.
Columns 2, 4, 6, and 8 include school-year dummies.
The constant term in the FE regressions represents the average of all observations' school-year fixed effects.
Columns 5 - 8 use only the 2002-2004 sample.

Table 8: The average caste-gap by quintile

School-attendance area:	5km-radius from village centroid							
Dependent variable:	Percent of total points							
Sample:	Full sample				Excluding metropolitan areas			
Medium:	English		Kannada		English		Kannada	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SC/ST	-0.073	-0.036	-0.050	-0.035	-0.095	-0.041	-0.052	-0.036
	(0.007)	(0.002)	(0.002)	(0.001)	(0.011)	(0.003)	(0.002)	(0.001)
Quintile 1	-0.323	–	-0.181	–	-0.296	–	-0.179	–
	(0.004)		(0.002)		(0.006)		(0.002)	
Quintile 2	-0.233	–	-0.139	–	-0.202	–	-0.137	–
	(0.003)		(0.002)		(0.005)		(0.002)	
Quintile 3	-0.159	–	-0.104	–	-0.126	–	-0.102	–
	(0.003)		(0.002)		(0.005)		(0.002)	
Quintile 4	-0.092	–	-0.062	–	-0.069	–	-0.060	–
	(0.003)		(0.002)		(0.004)		(0.002)	
SC/ST * Quintile 1	-0.001	0.014	-0.007	0.000	0.012	0.030	-0.005	0.001
	(0.010)	(0.003)	(0.002)	(0.001)	(0.017)	(0.005)	(0.002)	(0.001)
SC/ST * Quintile 2	0.014	0.015	-0.001	-0.002	0.008	0.024	0.001	-0.001
	(0.010)	(0.003)	(0.002)	(0.001)	(0.017)	(0.005)	(0.002)	(0.001)
SC/ST * Quintile 3	0.001	0.004	-0.001	-0.004	-0.011	0.011	0.000	-0.003
	(0.010)	(0.003)	(0.002)	(0.001)	(0.017)	(0.005)	(0.002)	(0.001)
SC/ST * Quintile 4	0.016	0.004	-0.001	-0.003	0.019	0.009	0.000	-0.003
	(0.009)	(0.003)	(0.002)	(0.001)	(0.015)	(0.004)	(0.002)	(0.001)
Constant	0.730	0.589	0.521	0.422	0.714	0.590	0.520	0.422
	(0.003)	(0.000)	(0.002)	(0.000)	(0.006)	(0.000)	(0.002)	(0.000)
P-value of Wald statistic:								
$\beta_{SC/ST} + \beta_{SC/ST*Quintile\ 1} > 0$		0.000		0.000		0.001		0.000
$\beta_{SC/ST} + \beta_{SC/ST*Quintile\ 2} > 0$		0.000		0.000		0.000		0.000
$\beta_{SC/ST} + \beta_{SC/ST*Quintile\ 3} > 0$		0.000		0.000		0.000		0.000
$\beta_{SC/ST} + \beta_{SC/ST*Quintile\ 4} > 0$		0.000		0.000		0.000		0.000
Observations	429,837	429,837	1,595,732	1,595,732	232,007	232,007	1,513,330	1,513,330

Standard errors in parentheses are clustered at the level of the school-year.
Columns 1, 3, 5, and 7 include a set of year dummies. Columns 2, 4, 6, and 8 include a set of school-year dummies.
Quintiles of schools' performance distribution are defined within school-attendance areas and are medium-specific.
An *area* is a circle with a 5km radius from the centroid of a village that contains a school.
A metropolitan area contains a village with a population that exceeds 1,000,000.
Results are estimated for the 2000-2004 sample only.

Table 9: The average caste-gap by quintile across management types

School-attendance area:	5km-radius from village centroid											
Dependent variable:	Percent of total points											
Sample:	Government				Aided				Unaided			
Medium:	English		Kannada		English		Kannada		English		Kannada	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
SC/ST	-0.044	-0.036	-0.039	-0.037	-0.102	-0.042	-0.059	-0.035	-0.064	-0.031	-0.040	-0.028
	(0.004)	(0.003)	(0.002)	(0.002)	(0.015)	(0.004)	(0.003)	(0.002)	(0.009)	(0.004)	(0.003)	(0.002)
SC/ST * Quintile 1	0.021	0.017	-0.011	0.006	-0.013	0.012	-0.005	-0.006	-0.019	0.014	-0.024	-0.010
	(0.006)	(0.005)	(0.003)	(0.002)	(0.019)	(0.005)	(0.004)	(0.002)	(0.015)	(0.008)	(0.004)	(0.003)
SC/ST * Quintile 2	0.014	0.019	-0.009	0.001	0.017	0.017	0.004	-0.004	0.002	0.005	-0.008	-0.006
	(0.006)	(0.005)	(0.003)	(0.002)	(0.020)	(0.005)	(0.003)	(0.002)	(0.013)	(0.006)	(0.005)	(0.004)
SC/ST * Quintile 3	0.012	0.011	-0.008	0.000	-0.009	-0.002	0.002	-0.005	0.005	0.007	-0.008	-0.008
	(0.005)	(0.005)	(0.003)	(0.002)	(0.020)	(0.005)	(0.003)	(0.002)	(0.013)	(0.005)	(0.005)	(0.004)
SC/ST * Quintile 4	0.001	-0.006	-0.007	-0.001	0.021	0.008	0.003	-0.005	0.015	0.010	-0.006	-0.003
	(0.005)	(0.005)	(0.003)	(0.002)	(0.020)	(0.005)	(0.003)	(0.002)	(0.012)	(0.005)	(0.005)	(0.003)
P-value of Wald statistic:												
$\beta_{SC/ST} + \beta_{SC/ST*Quintile\ 1} > 0$		0.000		0.000		0.000		0.000		0.004		0.000
$\beta_{SC/ST} + \beta_{SC/ST*Quintile\ 2} > 0$		0.000		0.000		0.000		0.000		0.000		0.000
$\beta_{SC/ST} + \beta_{SC/ST*Quintile\ 3} > 0$		0.000		0.000		0.000		0.000		0.000		0.000
$\beta_{SC/ST} + \beta_{SC/ST*Quintile\ 4} > 0$		0.000		0.000		0.000		0.000		0.000		0.000
Observations	154,972	154,972	792,154	792,154	131,531	131,531	621,954	621,954	143,322	143,322	181,624	181,624

Standard errors in parentheses are clustered at the level of the school-year.
Columns 1, 3, 5, 7, 9, and 11 include a set of year dummies. Columns 2, 4, 6, 8, 10, and 12 include a set of school-year dummies.
Quintiles of schools' performance distribution are defined within school-attendance areas.
Estimates of the quintile main effects in columns 1, 3, 5, 7, 9, and 11 have not been reported to save space.
An *area* is a circle with a 5km radius from the centroid of a village that contains a school.
Results are estimated for the 2000-2004 sample only.

Table 10: The caste-gap by percentile within English schools

Variable:	School-level differences in SC/ST and non-SC/ST percentiles											
Medium:	English											
Sample:	All schools			Top			Middle			Bottom		
Percentile:	10	50	90	10	50	90	10	50	90	10	50	90
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Panel A:	**All management types**											
Difference	-0.018	-0.037	-0.046	-0.040	-0.052	-0.053	-0.012	-0.035	-0.039	-0.018	-0.049	-0.070
	(0.005)	(0.007)	(0.005)	(0.018)	(0.018)	(0.007)	(0.006)	(0.008)	(0.006)	(0.021)	(0.016)	(0.012)
Observations	4324	4324	4324	777	777	777	2501	2501	2501	598	598	598
Panel B:	**Government schools**											
Difference	-0.003	-0.021	-0.049	-0.033	-0.054	-0.055	-0.009	-0.032	-0.051	0.004	0.001	-0.043
	(0.012)	(0.011)	(0.005)	(0.017)	(0.006)	(0.010)	(0.011)	(0.007)	(0.006)	(0.040)	(0.022)	(0.013)
Observations	960	960	960	127	127	127	534	534	534	179	179	179
Panel C:	**Aided schools**											
Difference	-0.021	-0.033	-0.044	-0.047	-0.046	-0.027	-0.029	-0.035	-0.039	-0.006	-0.027	-0.077
	(0.008)	(0.007)	(0.005)	(0.019)	(0.017)	(0.009)	(0.012)	(0.011)	(0.007)	(0.015)	(0.014)	(0.027)
Observations	718	718	718	129	129	129	386	386	386	126	126	126
Panel D:	**Unaided schools**											
Difference	-0.024	-0.034	-0.038	-0.031	-0.017	-0.042	-0.005	-0.03	-0.034	-0.014	-0.145	-0.038
	(0.009)	(0.007)	(0.005)	(0.012)	(0.012)	(0.006)	(0.009)	(0.009)	(0.009)	(0.023)	(0.024)	(0.027)
Observations	824	824	824	186	186	186	479	479	479	63	63	63

Standard errors in parentheses are clustered at the level of the village.
Results are estimated using only the 2000-2004 sample.
The difference is percentiles is measured as the SC/ST percentile in the school minus the non-SC/ST percentile.
Top schools are in the fifth quintile of their area's performance distribution; bottom schools, in the first quintile; and middle schools in the second, third, and fourth quintiles.

Table 11: The caste-gap by percentile within Kannada schools

Variable:	School-level differences in SC/ST and non-SC/ST percentiles											
Medium:	Kannada											
Sample:	All schools			Top			Middle			Bottom		
Percentile:	10	50	90	10	50	90	10	50	90	10	50	90
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Panel A:	**All management types**											
Difference	-0.03	-0.028	-0.045	-0.025	-0.021	-0.029	-0.035	-0.029	-0.047	-0.021	-0.029	-0.047
	(0.002)	(0.001)	(0.002)	(0.008)	(0.005)	(0.006)	(0.003)	(0.002)	(0.003)	(0.005)	(0.004)	(0.005)
Observations	33801	33801	33801	3691	3691	3691	17941	17941	17941	6573	6573	6573
Panel B:	**Government schools**											
Difference	-0.029	-0.029	-0.049	-0.042	-0.036	-0.049	-0.031	-0.03	-0.048	-0.021	-0.025	-0.053
	(0.002)	(0.001)	(0.002)	(0.012)	(0.007)	(0.010)	(0.003)	(0.002)	(0.003)	(0.004)	(0.003)	(0.005)
Observations	9535	9535	9535	799	799	799	4802	4802	4802	1862	1862	1862
Panel C:	**Aided schools**											
Difference	-0.032	-0.03	-0.043	-0.034	-0.03	-0.04	-0.036	-0.031	-0.047	-0.024	-0.027	-0.031
	(0.003)	(0.002)	(0.003)	(0.008)	(0.005)	(0.008)	(0.004)	(0.002)	(0.004)	(0.007)	(0.004)	(0.007)
Observations	6210	6210	6210	770	770	770	3555	3555	3555	935	935	935
Panel D:	**Unaided schools**											
Difference	-0.032	-0.018	-0.031	-0.004	-0.012	-0.007	-0.042	-0.019	-0.029	-0.025	-0.026	-0.039
	(0.009)	(0.004)	(0.006)	(0.023)	(0.011)	(0.016)	(0.013)	(0.005)	(0.008)	(0.017)	(0.012)	(0.013)
Observations	2015	2015	2015	315	315	315	992	992	992	467	467	467

Standard errors in parentheses are clustered at the level of the village.
Results are estimated using only the 2000-2004 sample.
The difference is percentiles is measured as the SC/ST percentile in the school minus the non-SC/ST percentile.
Top schools are in the fifth quintile of their area's performance distribution; bottom schools, in the first quintile; and middle schools in the second, third, and fourth quintiles.

Table 12: Variation with population density and an area's SC/ST share

Dependent variable:	School-level differences in SC/ST and non-SC/ST percentiles					
Sample:	All schools					
Medium	**English**			**Kannada**		
Percentile:	10	50	90	10	50	90
	(1)	(2)	(3)	(4)	(5)	(6)
SC/ST share	-0.036	-0.038	-0.038	0.007	-0.005	0.021
	(0.013)	(0.012)	(0.011)	(0.006)	(0.005)	(0.007)
$N \times 10^{-5}$	-0.015	-0.013	-0.005	0.016	0.013	0.021
	(0.004)	(0.004)	(0.003)	(0.005)	(0.001)	(0.002)
Constant	-0.004	-0.019	-0.038	-0.033	-0.028	-0.052
	(0.006)	(0.007)	(0.004)	(0.002)	(0.002)	(0.003)
Observations						

Standard errors in parentheses are clustered at the level of the village.
Only observations from the 2000-2004 sample are used.
The difference is percentiles is measured as the SC/ST percentile in the school minus the non-SC/ST percentile.
Top schools are in the fifth quintile of their area's performance distribution; bottom schools, in the first quintile; and middle schools in the second, third, and fourth quintiles.
The SC/ST share represents medium-specific shares of SC/ST test-takers in attendance areas defined by a 5km radius from the village centroid. N represents population density (individuals per square kilometer), and is measured from the 2001 Census of India.

Table 13: Evidence of ex-post ability sorting

Medium:	English		Kannada	
School count in area:	At most 10 schools	More than 10 schools	At most 10 schools	More than 10 schools
	(1)	(2)	(3)	(4)
rank(25th) = rank(50th)	0.808	0.760	0.690	0.791
	(0.005)	(0.003)	(0.002)	(0.002)
rank(50th) = rank(75th)	0.803	0.750	0.677	0.786
	(0.005)	(0.003)	(0.002)	(0.003)
rank(25th)=rank(75th)	0.709	0.649	0.585	0.707
	(0.006)	(0.004)	(0.002)	(0.003)
rank(25th) = rank(50th) = rank(75th)	0.675	0.595	0.519	0.663
	(0.006)	(0.004)	(0.002)	(0.004)
Observations	3,607	1,811	20,938	6,951

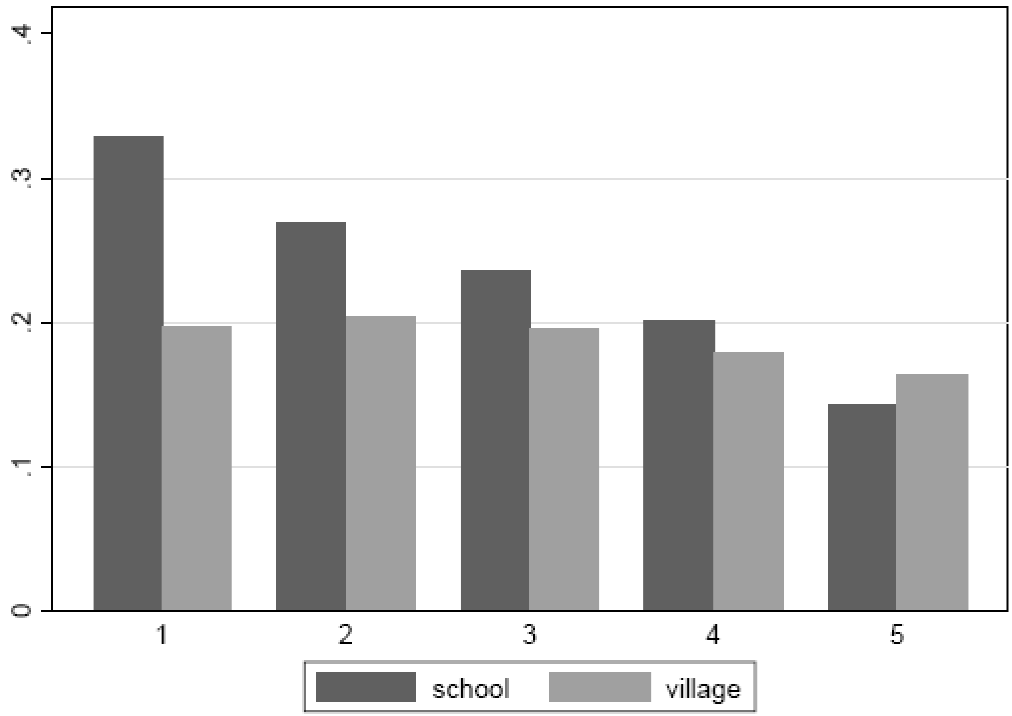
Observations are school attendance area - years in the 2000-2004 period.
School attendance areas are defined using 5km radii from village centroids.
Within each school attendance area-year, schools were ranked by their 25th, 50th, and 75th percentiles. In columns 1 and 3, a school was assigned a value of 1 *for 25th = 50th* if the rank of its 25th percentile was equal to the rank of its 50th percentile. In columns 2 and 4, it was assigned a value of 1 when the quintile of its 25th percentile was equal to the quintile of its 50th percentile.
The average of these rank (quintile) equivalence indicators was then taken across schools within an attendance area - year. Thus, column 1, for example, indicates that the average school-attendance area with 10 or less English schools in the 2000-2004 period had 71% of its schools retain the same rank within the area for their 25th and 75th percentiles of test scores.

Table 14: Age differences across castes

Dependent variable:	Age			
Medium:	English		Kannada	
	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)
SC/ST	0.139	0.061	0.185	0.171
	(0.005)	(0.003)	(0.002)	(0.002)
Constant	15.386	15.332	15.736	15.621
	(0.011)	(0.000)	(0.007)	(0.000)
Observations	855,144	855,144	3,479,215	3,479,215

Standard errors in parentheses are clustered at the school-year level.
Columns 1 and 3 include year dummies.
Columns 2 and 4 include school-year dummies.
The constant term in the FE regressions represents the average of all observations' school-year fixed effects.

Figure 1: SC/ST shares by school performance quintile



Sources: Test-score files of the March 2004 exam administered by the Karnataka Secondary Education Examination Board; Census of India, 2001.

School-performance quintiles were constructed by ranking schools according to the pass rates of their non-SC/ST students. The school share in a quintile represents the average SC/ST share of schools in that quintile. The village share in a quintile represents the average SC/ST share of the villages in which the schools in the quintile operate.

Figure 4: Imperfect screening and statistical discrimination

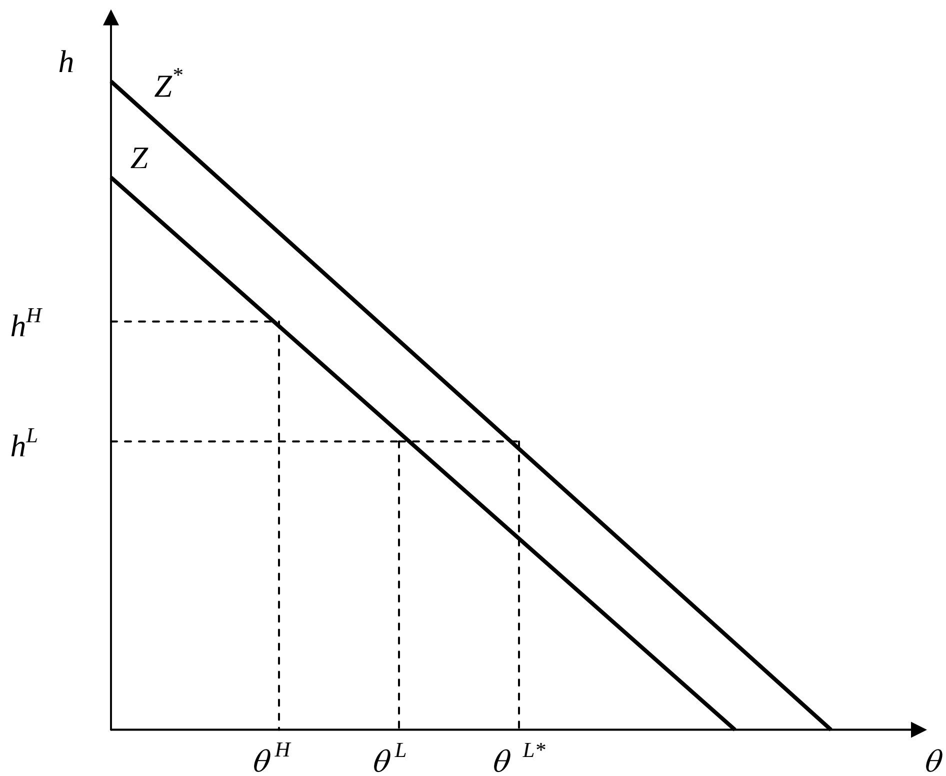
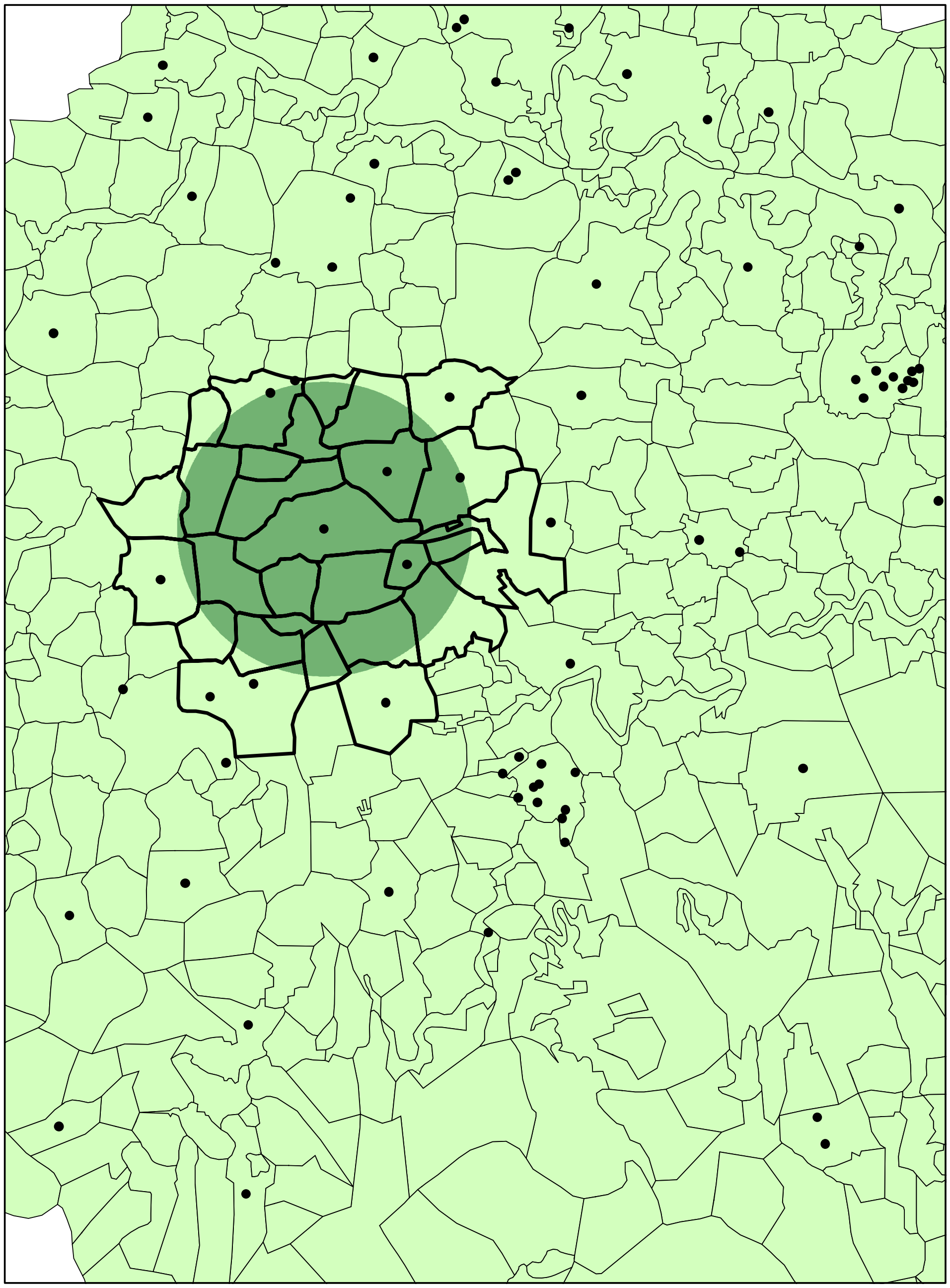
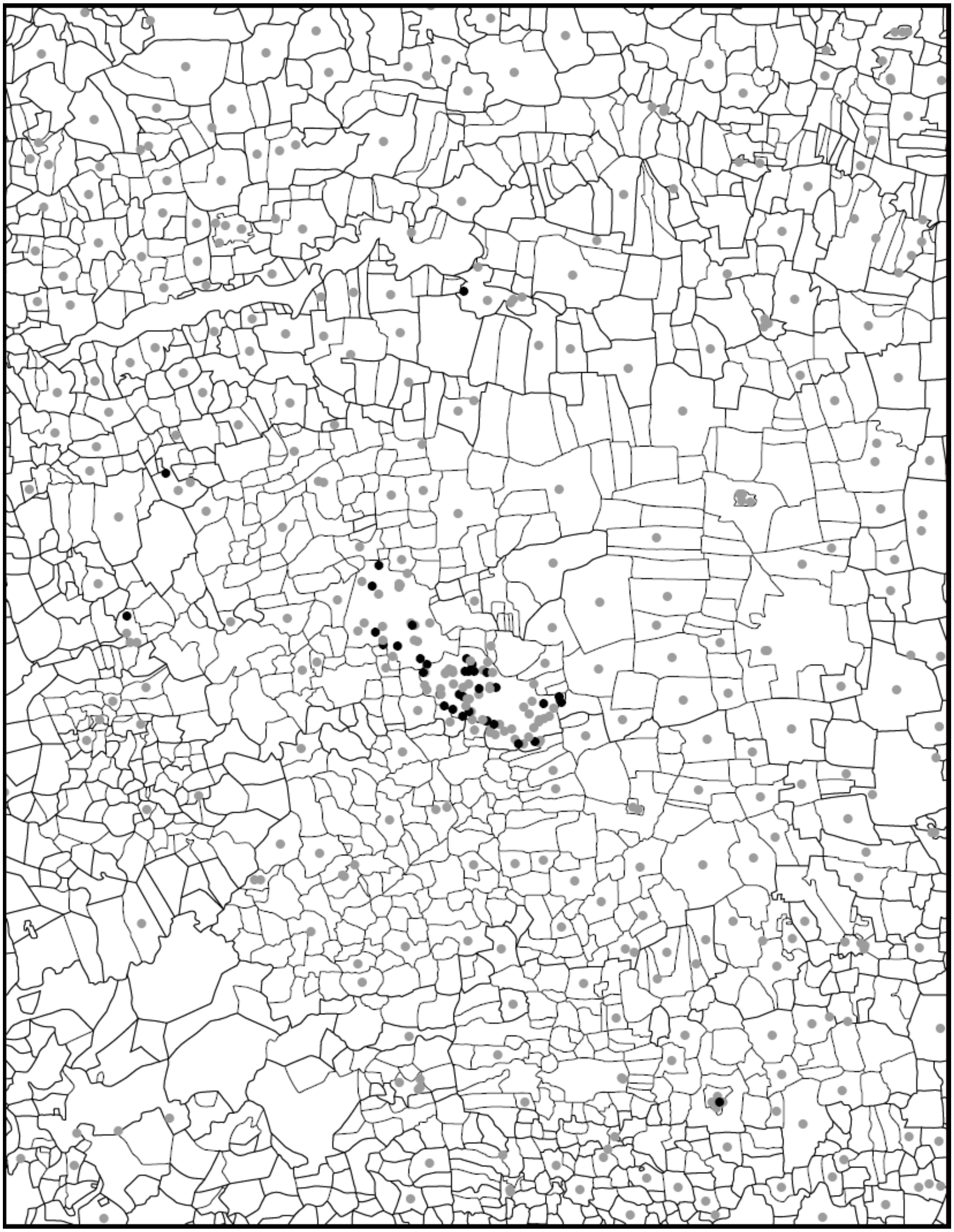


Figure 5: Construction of school-attendance areas



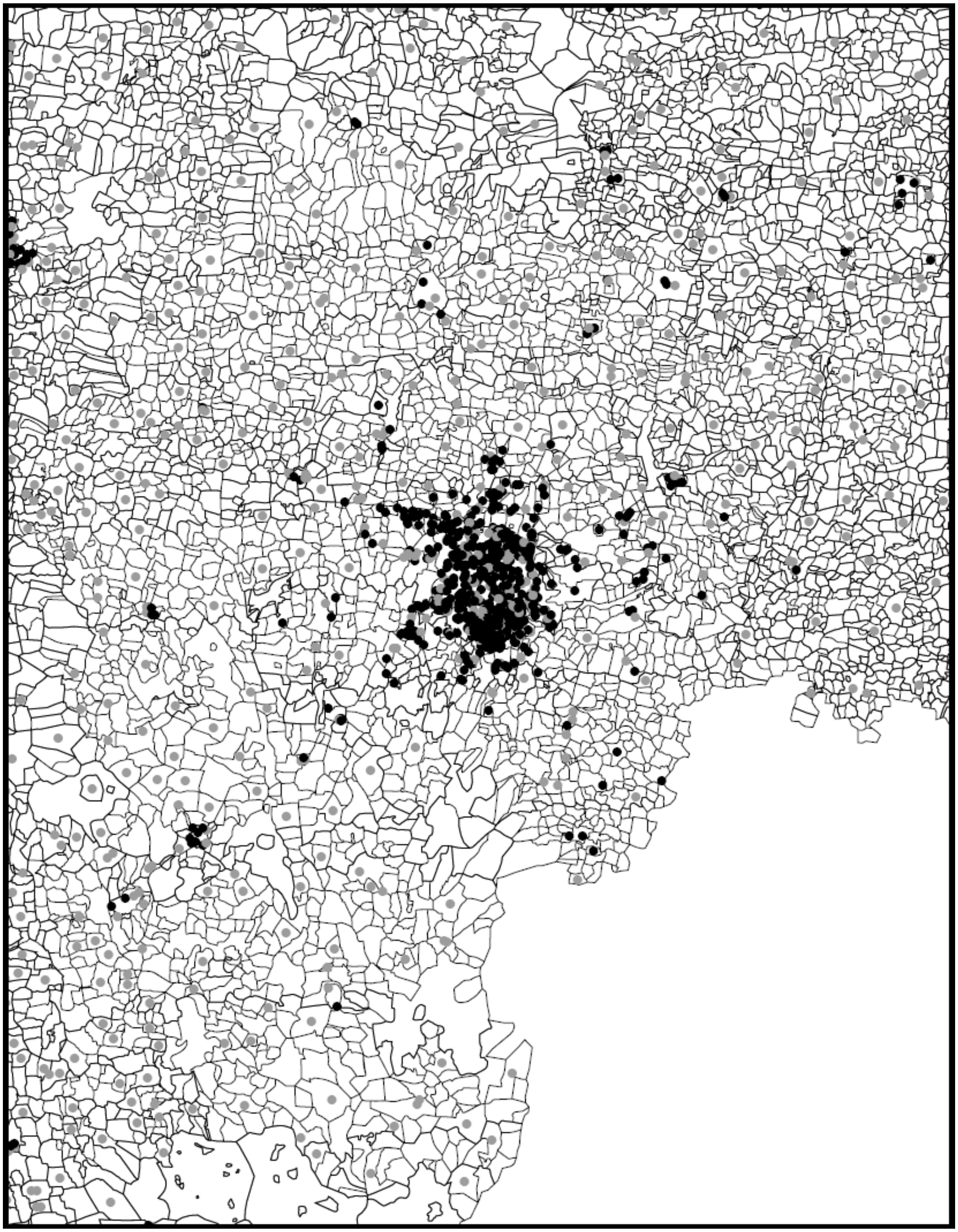
This map depicts the construction of a school-attendance area with a radius of 5km around Makodu village in the district of Mysore. The dots represent schools, and the darkly outlined villages all belong to the school-attendance area of Makodu, situated in the center of the circle.

Figure 6: The spatial distribution of English and Kannada schools (Mysore)



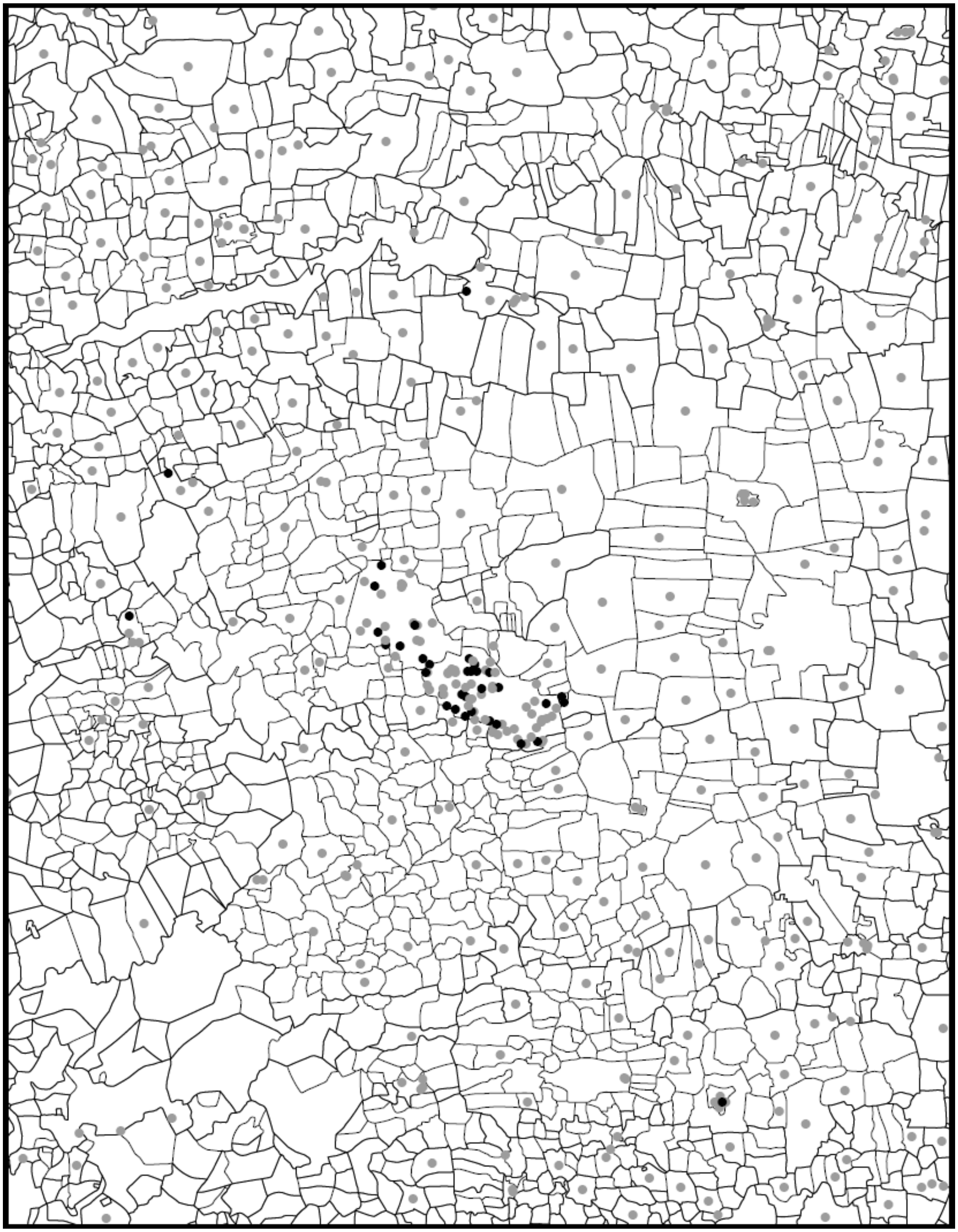
This map depicts English (black dots) and Kannada (gray dots) around the city of Mysore.

Figure 7: The spatial distribution of English and Kannada schools (Bangalore)



This map depicts English (black dots) and Kannada (gray dots) around the city of Bangalore.

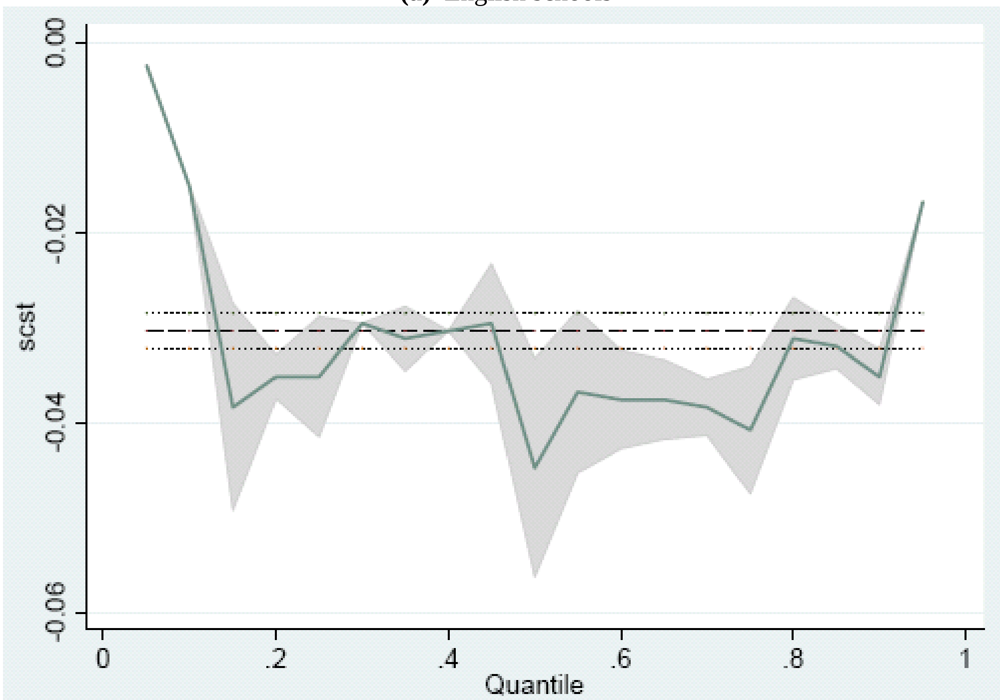
Figure 8: The spatial distribution of English and Kannada schools (Dharwad)



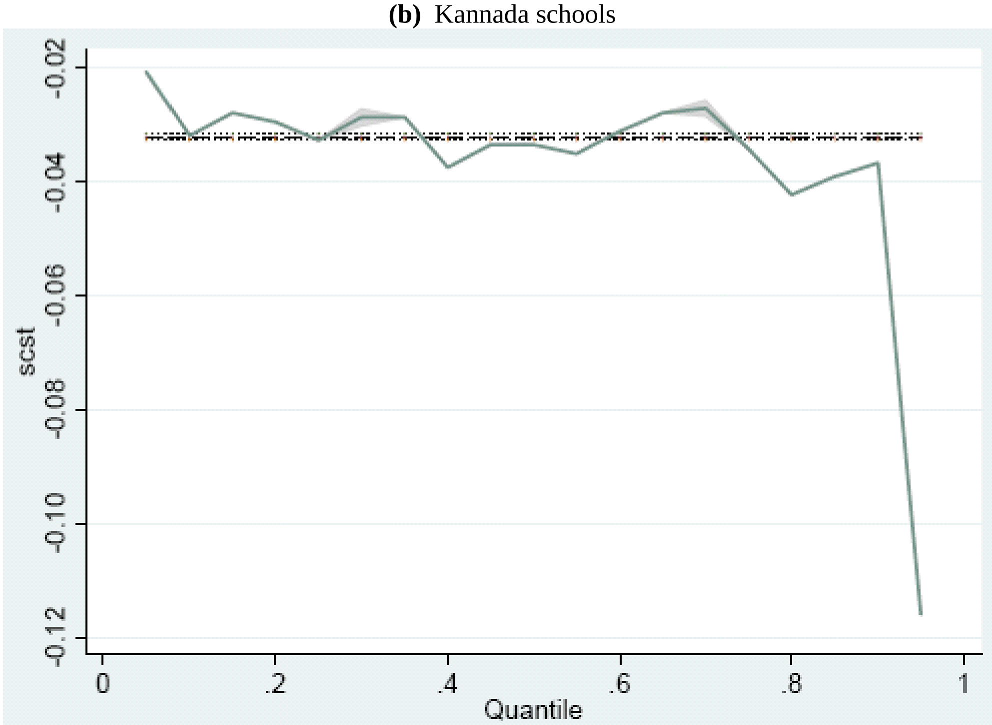
This map depicts English (black dots) and Kannada (gray dots) around the city of Hubli-Dharwad.

Figure 9: Quantile regressions

(a) English schools



(b) Kannada schools



Note: These regressions and those in Figure 9 employ only the 2000-2004 sample. The dashed and dotted lines represent the OLS estimates and their confidence intervals, respectively. The solid lines and grey shading represent the quantile estimates and their confidence intervals, respectively.

Figure 10: Quantile regressions

Top schools | Middle schools | Bottom schools

English

Kannada