INDIA POLICY FORUM 2007

DRAFT PAPER

Can Schooling Policies Affect Schooling Inequality? An Empirical Evaluation of School Location Policies in India

Anjini Kochar*

June 2007

Abstract

The Government of India has long made access to primary schools a priority. Its school location policies are built around the objective of ensuring a school either within a habitation, or within easy walking distance. In designing this policy, scant attention was paid to the fact that school location policies affect more than just access; they combine with the residential structure of an economy to determine many important aspects of school quality such as school size, and hence the number of teachers, and the composition of the student body. In India, the small size of habitations and their distance from each other implies that important aspects of quality, such as school size and student composition, are determined by location decisions, rendering it difficult to affect changes in quality. Supporting these hypotheses, the empirical analysis of this paper shows that habitation size determines the number of teachers and the availability of schools in scheduled caste and tribe (SC/ST) habitations, and that these in turn determine schooling attainment. The availability of schools in SC/ST habitations suggests caste-based segregation in schools and, supporting this interpretation, the effect of this variable increases the schooling of upper castes but reduces that of scheduled castes. Thus, school location policies, through their effect on school quality, imply that the benefits of school access differ across regions, but also across castes within any given region.

* Stanford University, Stanford, USA

THE BROOKINGS INSTITUTION

1. Introduction

The Government of India's schooling policies have long been guided by two primary objectives: To increase schooling attainment and to reduce schooling gaps, particularly those based on caste and gender. As in most developing economies which started with poor social infrastructure in rural areas, it was felt that a primary deterrent to schooling was the inadequate number of schools and the consequent distance between the average residence and a school. The Government therefore made the provision of a school within walking distance from each rural household a priority.

In implementing this policy, scant attention was paid to the fact that targeting access to schools as a primary objective may constrain the government in addressing other critical aspects of schools, particularly those related to school quality. This is because decisions regarding the placement of schools determine more than just access to schools; they combine with the residential structure of an economy to define the school community, and hence characteristics of schools known to affect schooling attainment. For example, if residential centers are geographically dispersed and of relatively small size, then the priority placed on access may imply a corresponding inability to efficiently choose school size: it will be determined by the size of the community. And, if the economy is characterized by a relatively high level of residential segregation, decisions regarding the location of schools across closely spaced communities will determine whether residential segregation implies a correspondingly high degree of schooling segregation. Under such conditions, policies to improve school quality cannot be discussed independently of school location policies.

The nature of residential communities in rural India makes this trade-off between access and quality likely. Rural India resides in habitations – distinct residential sub-divisions of a village – which vary in size but are, on average, fairly small. Habitations are generally organized along caste lines, so that the rural economy is characterized by a considerable degree of caste-based segregation with scheduled caste households frequently residing in separate, smaller, sub-habitations of the village. The stated policy objective of providing a school within easy walking distance of each household in conjunction with the geographic distance across habitations requires the Government to use the habitation as the basis for school mapping exercises.

In turn, the small size of the average habitation removes the possibility of optimally choosing school enrollment or size; it is dictated by the population size of the habitation. Reflecting this inefficiency, school size varies tremendously within a district, but even within a village, and forges a similar inefficient variation in critical schooling inputs such as the number of teachers. If these inputs affect schooling attainment, then the benefits of access to schools will vary across habitations, depending on its size. Moreover, the geographical distance across habitations implies that large villages commonly have more than one school, with each school serving a different habitation. In these villages, the organization of habitations along caste lines gets translated into a corresponding organization of schools along caste lines.

The empirical analysis of this paper starts by exploring the relationships between habitation size and the attributes of rural schools. The positive correlation between habitation size and school access suggested by India's school location policy is borne out in the data. Additionally, and consistent with the hypothesis that school location policies forge a relationship between residential characteristics and other attributes of schools, the data reveal that large habitations also have an advantage in the availability of teachers; districts with larger habitations have a smaller proportion of schools with two or less teachers. Moreover, districts with larger habitations are also more likely to have relatively large scheduled caste habitations and hence a greater probability of schools being located in scheduled caste habitations. This suggests greater schooling segregation in these districts.

The paper then proceeds to explore the effect of these school attributes on primary enrollments. I show that the probability of a school being located within a child's habitation of residence *does* increase enrollment. Extending this analysis, I find that small schools, with two or fewer teachers, are less likely to attract students. Further, the extent of caste-segregation in schools, measured by the probability of a school being located in a SC habitation, also affects enrollment. These effects vary by caste. Schooling segregation *reduces* enrollments by scheduled caste students, though the effect is not statistically significant at conventional levels of significance. However, it significantly *increases* enrollments by children of other caste households.

Taken together, the results of this paper confirm a relationship between habitation size and schooling outcomes, not just through the effect of size on school access, but also because it influences attributes such as teacher strength and the availability of schools in SC communities. This analysis thus confirms that school location policies affect critical determinants of school quality, and contribute to the significant observed variation in school quality across regions. They also contribute to caste-based schooling inequalities: Because regions characterized by larger habitations have an advantage in access, but are also more likely to be characterized by caste-segregation across schools, the benefits of improved access are magnified for members of upper castes, but reduced for scheduled caste households. Extending the results of this paper, I quantify the extent of this increase.

This paper is related to several literatures. Most closely related is the theoretical and empirical literature on the effect of community characteristics on schooling outcomes (Fernandez and Rogerson, 1996; Borjas 1995; Case and Katz 1991; Coleman 1988, Miguel and Gugerty 2005). As in this literature, I argue that community characteristics play an important role in shaping schooling attainment. However, the focus on school location policies suggests that habitation size may be as or more important than attributes such as community income and its ethnic composition which have been the focus of the literature to date. Because school size affects so many critical school inputs, this focus suggests that the influence of school location policies may be far-reaching: they can affect a very wide range of critical school inputs.

The primary contribution of my research to this literature is on the empirical side. The available empirical work on the topic documents the effect of community characteristics on schooling attainment, but generally does not show that these are mediated through the effect of the community on schooling inputs. Such a structural analysis has proven difficult, both because of lack of school-level data on school inputs and because of the lack of credible instruments to identify their effects. Using community characteristics to identify the effect of school inputs is questionable, since characteristics such as the ethnic composition of the community will likely affect schooling attainment directly, not just through school characteristics. Moreover, when a school serves several different residential communities, data is required not just on the community in which the school is located, but on all those that fall within the school's catchment area.

I use the rules which determine whether schools should be placed within a habitation in conjunction with habitation characteristics as the basis for identification of the effect of school characteristics. Doing so directly relates schooling outcomes to the combined influence of school location policies and habitation characteristics, and thus provides evidence on the role of school location policies in shaping schooling outcomes. I limit myself to an analysis of the effect of those characteristics which I can identify through this approach: access to schools, the number of teachers in the school, and the extent of caste-based school segregation.

The Government's school mapping exercises are conducted at the level of the district, based on data on habitations and schools collected through the All India Education Surveys (AIES). I use the same data, combined with household data on school enrollment from NSS surveys, for the empirical analysis of this paper. The focus of this paper on the consequences of school location policies thus limits the analysis to the identification of district-level variation in school inputs. However, this approach also implies several benefits. It removes the necessity for school-level data, provided the appropriate district-level counterparts can be constructed. The very rich district-level data on the demographic characteristics of habitations and the habitation-wise availability of schooling inputs contained in the AIES facilitate such an analysis.

This is true also for estimates of the effect of school segregation. Recognizing that residential segregation will translate into school segregation only if schools are provided in SC/ST habitations in addition to other caste habitations, I use data on the availability of schools in SC habitations within the district to proxy the extent of school segregation in the district. Doing so yields several benefits. First, it emphasizes the role of school location policies, since the variable I consider is the availability of schools in SC/ST communities, an outcome which is determined by these policies. Second, the aggregate approach of this paper thus automatically deals with the difference between the schooling and residential community which arises when a school's catchment area spans several communities. A disadvantage of this approach is that conclusions regarding the effects of school segregation rely on the existence of a correlation between the availability of schools in SC habitations and schooling segregation. While caste-based residential segregation is strongly suggestive of this correlation, I cannot confirm it without data on the student composition of individual schools – data which are not available.

The focus on school inputs as the pathway through which habitation characteristics affect schooling attainment yields a key insight which has not been sufficiently developed in the existing literature. Existing theories note that community effects on schooling will generate regional inequality in schooling attainment: children who reside in "good" communities will do consistently better than other children, of similar individual and household characteristics, who reside in "bad" communities, generating persistent schooling inequalities across communities. The focus on schooling inputs, in conjunction with the recognition that their effect on attainment can vary across households suggests an alternative form of inequality. Specifically, within any given residential community, the benefits of access to schools may be significantly higher for some households relative to others.

This research is also related to a relatively small literature which examines the determinants of schooling segregation. For example, Benabou (1996) argues that the positive effects of characteristics such as the average wealth of the community will cause schooling segregation, with richer households separating into exclusive schooling communities so as to increase the mean value of local income. In contrast, observed patterns of residential segregation in many developing economies, such as India, have not arisen as a consequence of community influences on schooling production, but have been a historical response to the desire for segregation in the use of public goods (Kochar 2006). In this context, I show that existing patterns of residential segregation generate segregation in schools, but only because of the school location policy which provides each habitation, with a population above a defined threshold level, with a school.

Also related is an extensive empirical literature which examines the effect of access, or distance to school, on schooling outcomes, primarily enrollments. This literature generally yields low responsiveness to distance to school (Filmer 2004). However, these results have been questioned on the grounds that they do not allow for the possibility that Governments' decisions regarding school placement are responsive to unobserved regional variables which may also directly affect schooling choices, including the returns to schooling.[1] Studies which have attempted to deal with this endogeneity, such as Foster and Rosenzweig (1996), have used questionable instruments, and, moreover, have restricted their analysis to the effect of the availability of a school within a village, not allowing for residential segregation and the possibility that the location of the school within the village may be far more important to households.[2] In addition to the difficulties associated with the endogeneity of school location, a major shortcoming of this literature, which it shares with the policy debate on the topic, is that it ignores the possibility that decisions regarding school location may combine with residential patterns to affect other attributes of schools known to influence schooling outcomes.

The rest of this paper is organized as follows. Section 2 describes India's school location policies, as well as policies relating to the hiring of teachers and

---

[1] Examples of such research include Birdsall (1985), Lillard and Willis (1994), Alderman et al (1993), Deolalikar (1997).

[2] Foster and Rosenzweig regress enrollment on the availability of a school in the village and on household measures of agricultural productivity, using a data set with observations on households in 1971 and, again, in 1982. They control for endogeneity by including household fixed effects and instrumenting (the change in) school availability, yields and yield growth by inherited land, farm equipment and buildings, consumer durables and first period levels of assets and primary schooling. While these instruments are undoubtedly correlated with the endogenous variables of the regression, they will also be correlated with other omitted variables, such as household savings, which are a component of the regression error term.

decentralized planning which are relevant to this paper. Section 3 turns to the study region, the state of Uttar Pradesh, and discusses residential and schooling patterns in the state, drawing on the AIES data but also on a survey conducted by the World Bank on living conditions in Uttar Pradesh and Bihar. Section 4 describes the data, while Section 5 discusses the empirical methodology of this paper. The main results are presented in Section 6. The last section concludes.

2. India's School Location and Other Schooling Policies

2.1 School location policy

At Independence, the Government of India inherited a very weak schooling system, with educational facilities available only for 40% of the children in the 6-11 year age group.[3] Many felt that this reflected an inadequate access to schools. Though the total number of primary schools in 1950-51 was 2,09,671, so that there was one school for approximately 212 children of that age group, this number did not adequately represent the distance of households to schools and hence the costs of schooling. At the time of the Second AIES (1965), only 38% of rural habitations had a primary section in them.

To ensure access, the Government determined that every rural habitation would be "served" by a school, in that a primary school would be available either in the habitation or within easy walking distance.[4] This lead to a rapid growth in the number of schools. By the Third Survey (1973), the number of habitations with primary sections in them increased to 44.3%, while 75.6% were served by primary sections either within the habitation or within a walking distance of 1 km. By 2002, 54% of rural habitations had a school within them, up from 50% in 1993. However, there remains considerable variation in access to schools across states. In Uttar Pradesh, for example, the percentage of habitations with a school located within the habitation was only 41% in 2002, and as low as 29% in 1993.

[3] Government of India, First Five Year Plan.
[4] For the first two surveys, this meant that a school should be available within 1 mile. As of the third survey, this was changed to 1 km.

The implementation of this policy required data on the number of habitations in rural India, and the percentage of these which were already served by schools. To provide this, the Government initiated the All India Education Surveys with the specific objective of collecting the data necessary for school construction policies. These surveys constitute the only data sets which use the habitation as the survey unit, and which provide information on the availability of schools and schooling inputs at the level of the habitation. The data collected in any particular survey provides the basis for the Government's school construction program in the period between that survey and the next survey. Thus, for example, information contained in the 5th All India educational Survey conducted in 1986 provided the basis for the Government's school construction program between 1986 and 1993, as well as the data which guided the allocation of teachers in this period.

The first AIES report, submitted in 1957, specified the guidelines for school construction, guidelines which maintain even today. Every habitation with a population of 500 or more should have a primary school located within the habitation. For habitations with a population ranging from 300-499, this population criterion was combined with a distance criterion: Schools should be located in habitations of this size range only if there were no primary schools (existing or proposed) within a walking distance of half a mile.

In determining the policy for the placement of schools, little thought was given to the fact that decisions regarding school location would also affect other characteristics of schools, notably school size and the socio-ethnic-economic composition of the student body. Instead, the Government made clear that it considered these decisions to be separable. The Second AIES survey states that its objective was to develop an approach for the location of schools, not to deal with issues such as the number of divisions or classes available or necessary in any given standard, or the optimum size of the school or class. It argued that these decisions came under the jurisdiction of the state level educational administration, in contrast to decisions regarding school availability which were under the authority of the Central Government.

By 1986, the number of primary schools had increased from 2,09,671 to 5,29,392. However, despite the significant expenditure on school construction, enrollments remained low, particularly for children from scheduled caste and tribe households. In rural Uttar Pradesh, for example, 1999 NSS survey data reveal that of SC/ST children between the ages of 6 and 11, only 48% were currently enrolled in primary school. In contrast, this percentage was 56% of children of other castes.

2.2 Teachers

Attention to the availability of teachers and to the student-teacher ratio started around the 1980s. The Steering Group on Education, Culture and Sports for the Seventh Five Year Plan (1985-1990) targeted a teacher pupil ratio of 1:40. The availability of teachers became a central focus with the 1986 National Policy on Education. This Policy introduced a scheme entitled Operation Blackboard to ensure the minimum essential facilities to all primary schools.

The introduction of Operation Blackboard in 1987-88 modified the rules for the allocation of teachers. Rather than be guided strictly on the basis of a targeted teacher-pupil ratio, the scheme specified a minimum number of teachers who should be provided in each primary school. The original scheme called for a minimum of two teachers. An extension of the scheme in the Eighth Plan called for a minimum of three teachers in all schools with an enrollment of 80 or more. This target was subsequently relaxed so as to ensure 3 teachers for all schools whose enrollment exceeded 100 students.

2.3 Decentralized planning

The Government of India has always stressed the need for decentralized planning in the context of schooling. In the early years of planning, however, this policy was implemented at the level of the state. The Sixth and Seventh Plans, for example, introduced state-specific targeting of central grants for schooling (under

Centrally Sponsored Schemes), making funds available on a priority basis to "Educationally Backward" states. [5]

The Eighth Plan, however, and the approach papers to the Plan, argued that inter-district variation in schooling was more pronounced than inter-state variation. The Working Group on Childhood and Elementary Education for the Eighth Plan therefore undertook a ranking of districts in terms of educational achievements. Arguing that there was no guarantee that funds targeted for Educationally Backward states would reach the backward districts within the state, it required all funds provided from central schemes to be targeted to districts (regardless of the state in which they were located) on the basis of this ranking, and in proportion to the degree of backwardness. The rank was based on a composite index, which gave equal weight to the following four parameters (based on data available in 1991): the literacy rate, the female literacy rate, the gross enrollment ratio for primary level, and this same ratio for females. This same index guided the allocation of funds through the Eight and Ninth Plans.

3. Residential patterns in Uttar Pradesh

In 2002, there were 96,014 villages in Uttar Pradesh, and more than double that number of habitations (201,606), so that, on average, each village had at least two habitations. Of the total number of habitations, 47,946 were classified as scheduled class or tribe habitations, so that SC/ST habitations accounted for 23.8% of total habitations. This percentage mirrored that of the percentage of scheduled caste and tribes in the state's population (23.5%). The total population residing in SC/ST habitations was 30,006,468. This population total amounts to 89% of the estimated population of scheduled castes and tribes in the state,[6] and suggests a very high degree of residential segregation by caste.

[5] Andhra Pradesh, Assam, Bihar, Jammu and Kashmir, Madhya Pradesh, Orissa, Rajasthan, Uttar Pradesh, West Bengal and Arunachal Pradesh.

[6] Of course, SC and ST habitations will also include members of other backward castes. Data on the exact caste composition of each habitation is not available.

Habitations are, in general, small. Figure 1 plots the distribution of habitations, by size group, for the years 1986 to 2002. The data reveal that, in 1993, fully 47% of habitations had a total population size of less than 300, while as many as 66% were less than 500 in size. SC and ST habitations are smaller: the percentage of SC/ST habitations with a population size of less than 300 is as high as 56.3% (figure 2). Thus, the average scheduled caste or tribe child is likely to reside in a habitation of smaller size than is a child from a general caste.

The policy of defining a school's catchment area by the habitation of its location, in conjunction with the small average size of habitations, implies that school size is also typically very small (figure 3). In 1993, the average school size in rural Uttar Pradesh was 138 students, with 37% of schools having a size of less than 100.[7] Because teacher allocations are determined by the total enrollment in the school, the small size of schools implies a correspondingly small number of teachers in each school (figure 4). The vast majority of schools in rural Uttar Pradesh have 2 or fewer teachers, implying that multi-grade teaching, which combines students of several different grades, is the norm.[8]

School-level data on enrollments and teachers in the 7th AIES reveal the considerable variation in school size (total enrollment), and hence in related characteristics such as the student-teacher ratio, even within a block (Figure 5, 6). This variation confirms that the policy of using habitations as the basis for school mapping exercise, in conjunction with the small size of habitations, yields school populations which are not efficiently determined – if they were, there would be little variation in school size within this rather small geographical unit. In fact, large villages with several habitations in them are also characterized by significant variation in the size of schools located within a village. This is strikingly revealed in Figure 7, where, for one village, I have plotted the

[7] A school size of 100 or less is generally considered to be unviable.

[8] The effects of multi-grade teaching have not credibly been established, and many believe that its effects may be positive. However, there is agreement that the benefits only obtain if it is restricted to two contiguous grades (for example, grades 1 and 2). In rural India, however, the assignment of teachers is based on class size, which is determined by the number of children of a particular age group in the habitation. It is thus common to see a large number of students in one class, and a small number in other classes. According, the assignment of teachers to classes reflects class size, not an optimal allocation.

distribution of school size across the 8 schools located in the village. School size in this village ranges from less than 100 to nearly 400.

Data on the proportion of habitations with schools in them (figure 5) reveal that over 80% of habitations with a population of more than 1000 have a primary school located within the habitation. Of habitations with a size of 500-1000, approximately half have a school located in them. This number falls off sharply after that: in 1993, 24% of habitations with a population size of 300-500 had a primary school located within the habitation, while the corresponding percentage for habitations of size less than 300 was only 6.9%.

The AIES does not provide school level data on the caste composition of the student body[9]. Lacking this, it is not possible to show the extent of caste segregation in schools in Uttar Pradesh. However, the extensive residential segregation by caste, in separate habitations, suggests a correspondingly high segregation of students by caste. This situation is not unique to Uttar Pradesh; since the school location policy is one adopted by the Central government, the same pattern of caste segregation across schools exists in other parts of India. Data from a survey of 60 schools in rural Andhra Pradesh (45 “main village” schools and 15 schools located in SC hamlets) reveals the considerable caste segregation in this state.[10] In hamlet schools, the proportion of scheduled caste students to total students is as high as 66% (2004). Indeed, scheduled castes and Other Backward Castes comprise almost all the student body (92%). In striking contrast, scheduled castes comprise only 16% of the student body in “main village” schools. This survey also reveals the difference in size and correspondingly in teacher strength across hamlet and main village schools. The average (total) enrollment in hamlet schools in 2004 was 90 students, while the average number of teachers was 3. Conversely, main village schools had, on average, an enrollment of 177 students, with an average of 6 teachers.

---

[9] For the 7th survey, data on enrollment is available at the level of individual schools, but no details on the caste composition of students are available.

[10] The first round of this survey was conducted in 2002, by the author, in collaboration with the Byrraju Foundation, Hyderabad.

4. Data

For the analysis of this paper, I combine household data from the NSS on the schooling status of individuals, with district level information on school availability and habitation characteristics from the All India Education Surveys. Specifically, I examine the enrollment status of two cohorts of 6-11 year olds, from the 1993 and 1999 NSS employment surveys (round 50 and 55), and match this with data from the 6th (1993) and 7th (2002) AIES survey on the proportion of habitations within a district with schools in them.

Each survey round of the AIES provides the basis for the school construction and mapping programmes in the inter-survey years. Thus, the construction of schools between 1986 and 1993, and hence the availability of schools in 1993, is based on the data contained in the 5th AIES survey (1986), while the availability of schools in 2002 is based on the 6th (1993) survey. As explained fully in the next section, data from the preceding AIES survey thus provides the source of identification of the current availability of schools.

Matching of the NSS surveys and the AIES data is done at the level of the district. This exercise is complicated by the division of districts across survey of rounds and the creation of new districts. Because we use data spanning the period 1986 to 2002, matching is done on the basis of the 1986 classification of districts. Information regarding the division of districts, which enables the tracking of districts over time, is provided in the Census of India.

5. Habitation size and school attributes

This section examines simple correlations between the characteristics of habitations and schools, focusing on habitation size. The objective is to show a relationship between habitation attributes and school quality, as measured by the number of teachers and other features of schools. The next section provides evidence on the effect of these school features on enrollments, and allows an assessment of the extent to which the correlations between habitation size and schooling outcomes reflect habitation effects on school attributes.

In addition to school access, measured by the proportion of habitations in a district with a school located within the habitation, I examine the determinants of two other school characteristics: the proportion of schools with two or less teachers, and the probability of a school being located in a scheduled caste or tribe habitation (given by the ratio of the number of SC/ST habitations with schools to total habitations). This latter variable indicates the degree of caste segregation in schools, since it is only when a school is available in a scheduled caste habitation within the village that children of scheduled castes and tribes are likely to attend separate schools from children of other castes. The two characteristics of habitations which I initially focus are the proportion of habitations with a population size of 500 or more, and the average size of habitations in a district. The former variable is of particular interest, since it forms the basis of school location policies. The data are at the level of the district. School attributes are for from the 7th and 6th AIES (2002 and 1993 respectively), and are related to habitation characteristics from the prior round of the AIES (6th and 5th respectively, for 1993 and 1986).

Table 1 presents simple correlation coefficients for these variables. The data reveal that districts characterized by relatively large habitations are more likely to have a school located within them. They are, however, also more likely to have schools located within scheduled caste habitations, suggesting a greater degree of caste segregation in schools. Finally, larger habitations are associated with fewer schools with two or less teachers.

Table 2 elaborates on this evidence through OLS regressions which control for the effect of various other demographic and economic characteristics of habitations, including the average population and squared population of habitations in the district, the average population (and its square) of SC habitations, the number of habitations and squared habitations separately for total habitations and SC/ST habitations, the district female literacy rate, average per capita expenditure and education rank. In addition to the proportion of total and

SC habitations of size greater than 500, regressions include the proportion of total habitation and SC habitations with a population of 300-499, since school location policies also call for schools to be located in habitations of this size, if no school is available within a 1 km. distance. As before, all regressions are run on district level data. Because outcomes for any given district are likely to be correlated, all standard errors are clustered at the level of the district.

Supporting the evidence from simple correlation coefficients, the proportion of habitations of size 500 and more is positively and significantly associated with the proportion of habitations with schools, even after controlling for the average size of habitations (and its square). This suggests that policy rules *do* affect school location, since the rule dictates that particular population cut-off levels musts determine school placement. Interestingly, this same variable does not affect the availability of schools in SC/ST habitations – access to schools in these habitations reflects the size distribution of SC/ST habitations, as measured by the proportion of SC habitations of population between 300 and 499 and greater than 500. The population cut-off values that determine school placement also do not appear to affect teacher allocations: proportion of habitations of size between 300 and 499 and greater than 500 is not a significant determinant of the proportion of schools with 2 or less teachers. This provides supportive evidence that these population cut-off values *do* reflect policy rules; if they merely reflected a non-linear effect of habitation population size, they would also likely affect other schooling inputs. Teacher strength *is* significantly affected by the average population size of habitations: as expected, the proportion of schools with two or less teachers falls as population size increases.

6. **Empirical methodology for assessing the effect of school characteristics on enrollments**

6.1 Basic equation for examining the effect of access to school

The basic estimating equation for assessing the effect of access to schools on the enrollment status of child i in district j is the following:

$$(1) \qquad \Pr(enroll)_{ij} = \alpha_0 + \alpha_1 (Sch\_hab)_j + X_{ij}{}' \alpha_2 + Z_j{}' \alpha_3 + u_{ij}$$

The dependent variable in this regression, $\Pr(\text{enroll})_{ij}$, is an indicator variable which takes the value 1 if the child is currently enrolled in primary school, 0 otherwise. The regressor of interest, Sch_hab_j is the proportion of habitations in the district with a school located within the habitation. X_{ij} is a vector of child and household specific characteristics which determine school enrollment. This includes the age and squared age of the child, his or her gender, a dummy variable which records whether he or she is a member of a scheduled caste, household per capita expenditure, the demographic composition of the household as reflected in the division of household members across 10 sex-age groups, the maximum years of education of adult household males in two different age group (20-40 and 40-60), and an indicator variable for whether the household head has completed primary education.. Z is a vector of district variables, including the district average per capita expenditure, female literacy rates, the number of habitations and the proportion of scheduled castes and tribes to total population in the district

A primary concern of this paper is to examine the differential effects of school location policies on members of scheduled castes and tribes. Therefore, all regressions are run on a pooled sample of children in the 6-11 age group, but then also separately for children who are members of scheduled castes and tribes, and those who are members of general castes.

6.2 Identification of school access

The critical issue in the estimation of (1) relates to the identification of the effect of school availability on enrollment. Decisions relating to the location of schools are based on the size of the habitation. But, habitation size will also determine total enrollments in the school and hence the number of teachers and a number of other school attributes. Large habitations may also have better infrastructure such as roads which are also likely to increase the returns to schooling. If so, α_1 in equation (1) will be biased upwards.

I base identification on the program rules which specify a specific population cut-off level for decisions regarding school location. Specifically, only habitations with a population exceeding 300 are eligible for schools. However, the criterion for determining the eligibility of habitations of sizes 300-499 differs from those of lager habitations, in that it includes a distance criterion (no other school within a 1 km. distance) as well as a population criterion. Thus, population cut-off levels of 300 and 500 will differentially affect school access.

These population cut-off levels are unique to decisions regarding school location – they do not influence decisions regarding the availability of any other local public good, including health centres, roads or investments in sanitation. This suggests the feasibility of basing identification on this particular size classification of habitations in a district, specifically on the proportion of habitations of size greater than 300 and greater than 500. Because the AIES is conducted specifically for the purpose of school mapping, it provides information on the number of habitations in a district which fall into these population groupings.

As previously described, data from a specific survey year formed the basis for the school mapping exercises conducted in the inter-survey period between that and the next survey. Thus, identification of the availability of schools in 1993, for example, reflects data on the distribution of habitations in 1986. It is these lagged population figures, the exact data which guided school mapping exercises, which are used in this analysis.

Habitation size, however, also determines school size. Moreover, data on the availability of schools in habitations, summarized in figure 8, suggest that targets for school construction were not achieved – not all habitations of population size greater than 500 have a school located within them. This reflects the financial constraints faced both by the central government and the state government – the necessary funds required to provide the desired number of schools have not been available. Instead, governments have had to prioritize amongst habitations in determining school location, making decisions on the basis of the district's educational ranking, as discussed in Section 2 of this paper.

The availability of a school in a habitation thus reflects the combined effect of the population criterion and the district's educational rank. This suggests the use of the interaction of the proportion of habitations of size 300-499 (*hab_3)* and greater than 500 *(hab_5)* with the district's education rank *(ed_rank)* as instruments. This educational index provides a particularly good source of identification, since it ranks *all* districts in the country based on 1991 data, not just districts within a given state, and because the particular *rank* of any given district was meaningful – the division of central funds for schooling was to be in proportion to the specific cardinal rank of the district. Moreover, the district's education rank amongst all districts in the country was not used to guide any state or local government decisions regarding schooling investments.[11]

To ensure that identification comes only from the interaction of the population criterion with the district's education rank, *hab_3* and *hab*_5, as well as *ed_rank*, are included amongst the set of regressors, and hence excluded from the instrument set. I also include the average population size of habitations in the district amongst the regressors, as further assurance that identification does not come from habitation size.

6.3 Assessing the implications of school size

As previously noted, the decision to locate a school within a habitation suggests that the socio-economic features of the habitation, including habitation size, will affect schooling outcomes through their effects on school attributes such as school size and the composition of the student body. This section discusses the methodology used to evaluate how size affects enrollments.

Rather than consider the effect of school size, I assess the impact of the number of teachers in the school, an attribute which closely reflects school size. School size will primarily be determined by the population size of the habitation. However, habitation size will also affect schooling outcomes through its influence on other schooling inputs such as the number of teachers. Moreover, it

[11] This ranking is available in Annexure V of the Report of the Working Group on Early Childhood Education and Elementary Education Set up for Formulation of the Eighth Five Year Plan.

will also determine the availability of other local public goods, such as health centers, which may have spillover effects on schooling. Lack of good instruments for school size therefore dictates my decision to consider, instead, the number of teachers.

The number of teachers assigned to a school reflects several criteria discussed in Section 2 of this paper. One criterion is total school enrollment. However, this rule does not generate a good source of identification. It implies that teacher availability is determined by habitation which, as previously argued, is likely to be directly correlated with schooling outcomes for a variety of reasons. Instead, I base identification on the auxiliary set of rules specified under Operation Blackboard and the extension of this program in 1992, which determine the minimum number of teachers in a school. Under the expanded Operation Blackboard, each school with an enrollment of over 100 was to have 3 teachers (with funding provided by the Central Government). This implies a distinct population cut-off which determines whether the school would have two or fewer teachers: Two (or fewer) teacher schools would be those with a total enrollment of less than 100.

Identifying the number of teachers on the basis of this strategy requires information on the proportion of schools in a district with an enrollment of less than 100. Unfortunately, that information is not available in the AIES. However, the information necessary to predict this number is. Specifically, we assume that the distribution of school size follows a normal distribution, so that the proportion of schools of size less than 100 is:

$$(2) \qquad \Pr(Size <= 100) = \int_0^{100} \frac{1}{\sqrt{2\pi}\sigma} e^{-(100-\mu)^2/2\sigma^2}$$

In all probability the variation in school size will vary with the distribution of habitation size. I therefore assume that $\sigma=f(hab\_3)$. A linear approximation of this, inserted into (2), implies that the probability of school size being less than 100 is a function of $Size\_100$ = - [(100-μ)/(hab_3)], where μ is the (lagged) average enrollment in the school.

I use $(Size\_100)^2$ and $(Size\_100)^3$ as instruments for the proportion of schools with two or less teachers. Identification thus comes from a non-linear function of the cut-off enrollment size which determines this distribution of teachers (100), lagged average enrollment in the district, and the proportion of habitations of size 300-500 in the district, with the particular functional form suggested by the probability distribution of school size. Lagged average enrollments and habitation size are from data in the survey round of the AIES previous to the NSS survey round from which the current enrollment status of the student is drawn. That is, for children in the 6-11 age group in 1993, lagged average enrollments and habitation size are from the 5th (1986) round of the AIES.

As before, stronger identification comes from interacting $(Size\_100)^2$ and $(Size\_100)^3$ with the district's educational index rank. Because teachers provided under Operation Blackboard and the expanded version of this scheme were funded by the Central Government, districts which had the highest rank were most likely to receive these funds on a priority basis. The interacted variable, $(size\_100)^2*ed\_rank$ and $(size\_100)^3*ed\ rank$ ensures that identification does not just come from a non-linear function of (lagged) average enrollments and habitation size.

6.4 Assessing the implications of residential segregation

A final concern is that the decision to use the habitation as the basis for school mapping exercises forges a correlation between the socio-economic composition of the habitation and that of the student body. Because habitations are organized along caste lines, this system of residential segregation will then get translated into a *de facto* system of caste-based schooling segregation. But, school segregation will only exist if schools are placed in scheduled caste habitations. If not, then scheduled caste students must attend the same schools as students of other castes. This suggests that the extent of schooling segregation will reflect the availability of schools in scheduled caste habitations.

I therefore examine the effect of the probability of a school being located in a scheduled caste habitation on enrollment decisions. This probability is given by the joint probability of a school being located in a habitation and the habitation being a scheduled caste habitation, whose sample counterpart is the ratio of the number of scheduled caste habitations with a school to the total number of habitations in the district. I refer to this variable as *(Sch_SChab)*.

The analysis of the effect of the availability of schools in scheduled caste habitations, as distinct from that of overall access to schools, is possible because the AIES provides data on the total number of scheduled caste habitations, their population distribution, and the number that have a school located within them. As in the treatment of the endogeneity of the (general) availability of schools *(Sch_hab)*, my instruments for *(Sch_SChab)* are formed by interacting the district's educational rank with the number of SC habitations of size 300-499 and the proportion of OSC habitations with a population of 500 or more. These two population characteristics of SC habitations, *SChab_3 and SChab_5* are included amongst the regressor set. I also expand the regressor set to include the proportion of SC habitations in the district, and the average size of SC habitations.

The expanded regression that I run is:

$$
(2) \qquad \Pr(enroll)_{ij} = \alpha_0 + \alpha_1 (Sch\_hab)_j + X_{ij}'\alpha_2 + Z_j'\alpha_3 + \alpha_4 (Tch\_2)_j + \alpha_5 (Sch\_SChab)_j + u_{ij}
$$

where $(Tch\_2)_j$ is the district proportion of schools with two or less teachers.

The first-stage reduced form regressions for *(Tch_2)* and *(Sch_SChab)* include the vector of variables which capture the population distribution of habitations in the district (*hab_3, hab_5, SChab_3, SChab_5)* as well as the interaction of these variables with the district's educational rank. Thus, the first stage regressions provide evidence of the effect of the size of the habitation on schooling outcomes. Combining this with the second-stage estimates of the effect of school attributes on enrollment outcomes (the coefficients in equation (2)), it is possible to examine the extent to which habitation characteristics affect schooling outcomes through their effect on school characteristics, independent of any direct effect through other means.

7. Results

7.1 Preliminary Regressions on School Access

Table 3 reports results from an initial set of regressions, which considers the effect of the availability of a school in the habitation on enrollment decisions (equation 1). The first column reports the reduced form first-stage regression of the proportion of habitations with schools in them on the set of instruments. All regression errors are clustered at the district level, to account for the correlation between observations within a given district. The second column reports results from a probit regression of school enrollment on *(Sch_hab)*, ignoring the endogeneity of this variable. The third column uses an instrumental-variable probit (IV probit) regression, with suitably adjusted standard errors, to account for endogeneity, using the instruments previously described. The last column lists the marginal effects of the variable on the probability of current enrollment in primary school, based on the IV probit results of the previous column.

The first stage regression attests to the explanatory power of the instruments. Larger habitations (with population in excess of 500) are more likely to have a school in them. However, the effect of population size on school access is tempered by the district's educational rank in the expected direction. Government policy specified that districts with a *lower* educational rank should first be allotted schools. Reflecting this, the interaction of habitation size with educational rank negatively affects the proportion of habitations in the district with a school. That is, a district with sufficiently large habitations was less likely to get a school if its educational rank was high.

The coefficient on school access obtained from probit regressions, which ignore the endogeneity of school location policies, is significantly lower than that obtained from IV probit regressions, which utlize the interaction of the variables used to guide policy with the education rank of the district as instruments. Correspondingly, Wald tests for the exogeneity of school access reject this hypothesis. These results suggest that prior estimates of a small effect of distance to school on enrollment decisions may be biased downwards, because of the failure to properly account for the endogeneity of school access. The IV probit results suggest a relatively large effect of the availability of schools within a habitation. A 1% increase in the proportion of habitations with a school will increase the proportion of school-age children attending schools by 0.4%. This justifies the decision to place a school within easy access to students, if this policy is to be evaluated purely from the viewpoint of its effect on school access.

Table 4 provides estimates of school access by caste and by gender. While school access is important for children from all castes, and for boys and girls, it matters more for scheduled castes and for girls. Thus, the elasticity of enrollment with respect to school access if as high as 0.7 for scheduled caste and tribe children, but only 0.3 for members of other castes. Similarly, this elasticity is 0.3 for boys, but 0.5 for girls.

7.2 Incorporating teacher availability and school segregation

I now extend the analysis to see how enrollment decisions are affected by additional school characteristics, and to examine whether these effects vary by caste. Table 5 documents results from the first stage regressions of the endogenous variables on the full set of instruments and additional regressors.

The regression results confirm the explanatory power of the instruments, and hence the role of habitation size in determining important school attributes. For example, the interaction of a district's education rank with the proportion of habitations of size greater than 500, significantly affects the proportion of habitations with a school. And, the interaction of $size\_100)^2$ and $(size\_100)^3$ with education rank also significantly reduces the proportion of schools with less than two teachers: As prescribed by policy, a greater proportion of schools with enrollment over 100 suggested additional teachers, and hence a reduction in the proportion of schools with two or fewer teachers at the time of the next survey. However, this effect is stronger in districts with lower educational rank.

Confirming the results from simple correlation coefficients, the first stage regressions also suggest that the instruments for the proportion of schools with two or less teachers do not also affect access to schools. This supports the credibility of the results, since it suggests that the instruments for teacher strength are uncorrelated with the availability of schools and other schooling inputs.

These first stage regression results provide the basis for the IV probit regressions in table 6. The first column repeats regression results for the probability of current enrollment in primary schools as a function of school access, ignoring other school attributes, to provide a point of comparison.[12] The regression reported in column (2) includes the proportion of schools with 2 or less teachers, and the probability of a school in a SC/ST habitation amongst the regressors. The results confirm that school access increases enrollment. However, as expected, fewer teachers imply a reduction in enrollment. And, higher levels of school segregation, suggested by the availability of schools in SC habitations, appear to *increase* enrollments.

Together in conjunction with the first stage regression results of table 5, these results therefore suggest that habitation size *does* affect schooling outcomes because of school location policies which determine whether a school can be constructed in a particular habitation, including the availability of schools in SC/ST habitations. Habitation size also affects school outcomes through its affect on the number of teachers in a school, which policy dictates must also be determined by school size. That the effects of habitation size on schooling outcomes are mediated through school inputs is further suggested by comparing the results in column 1, which conditions only on school access, with those in column 2 which additionally control for teacher strength and the provision of schools in SC/ST communities. The inclusion of these latter variables

[12] This regression differs from that reported in table 3, since it includes additional regressors.

significantly affects the coefficient on measures of habitation size which are included as regressors in both regressions.

7.3 Robustnss checks

Before assessing the separate implications of schooling inputs by caste, I conduct several robustness checks. First, I include $(size\_100)^2$ and $(size\_100)^3$ amongst the regressors, allowing identification of teacher strength to come only from the interaction between these variables and the district's education rank. This is a simply over-identification test for the validity of including these variables in the instrument set. The results, detailed in the last column of table 6, support their validity: their independent effect on school enrollment is statistically insignificant, and does not change the regression results.

A second check explores the sensitivity of results to non-linearity in habitation size. As always, when identification rests on a particular functional form assumption, there is the possibility that the results simply reflect a non-linear relationship between habitation size and school enrollment. To test for this, a second set of regressions includes the square of average habitation population as well as the square of the average population of SC habitations in the regressor set. The results, reported in table 7, suggest that the results are robust to this inclusion. The estimates of the effect of school inputs, including school access, are not affected by this inclusion. And, standard F tests for the expanded regression reject this version in favour of the more parsimonious specification.

7.4 Regression results by caste

In order to examine whether the effects of school attributes varies by caste, I run the same regression on schooling enrollments separately for children from scheduled castes and tribes and for those from other castes. The results are reported in table 8. For all children, regardless of caste, access to schools enhances the probability of enrollment, while the proportion of schools with two or fewer teacher reduces this probability. However, there are significant differences in the effect of schooling segregation. An increase in the probability of a school being located in a SC/ST habitation significantly enhances the schooling of upper caste households. Conversely, the effect on children from scheduled castes and tribes is negative. However, this latter effect, though relatively large in magnitude, is imprecisely estimated and is not significant at conventional levels of significance.

The results suggest that a greater degree of homogeneity in the student body enhances the schooling attainment of upper caste households. It is beyond the scope of this paper to explore the reasons which underlie this effect. One explanation offered in the literature is that ethnic homogeneity fosters community support for, and involvement in, schools, thereby improving schooling attainment (Miguel and Gugerty 2005). However, if this were so, the greater ethnic homogeneity fostered by schooling segregation should increase schooling enrollments for members of scheduled castes as well as those of other castes. A more convincing explanation is that schooling attainment does depend on the "mean" quality of the community, either through its influence on the attributes of the student body or through the quality of community contributions to the schools. If so, and if scheduled caste communities are characterized by lower mean quality then, consistent with the results of this paper, schooling segregation will enhance the schooling of upper caste households while reducing it for scheduled castes and tribes.

7.5 Quantifying the differential effects of habitation size on enrollment through school characteristics

Because the decision to locate a school in a habitation depends on its size, it is natural to expect that habitation size will positively affect schooling attainment, through its effect on access to schools. The empirical analysis of this paper shows, however, that habitation size also affects other school attributes, specifically teacher strength and the extent of schooling segregation. That is, the choice to locate a school in a habitation of a particular size determines access, but also attributes which determine the quality of the school. How do these auxiliary effects on school quality change the relationship between habitation size and enrollment that would result if school location policies did not affect other school attributes?

The answer to this question can be obtained by combining the first stage regression results on the determinants of school attributes with the IV probit estimates of the effect of school inputs on enrollment. However, because the first stage regressions utilize a number of different measures of habitation size, the overall effect of habitation size on school attributes are difficult to infer from the regression results reported in table 5.

I therefore proceed by first running a reduced form predicting equation for each of the school attributes in question (proportion of habitations with a school located in them, the proportion of schools with two or less teachers, and the probability of a school in a SC/St habitation) on average habitation size and its square. I use these regressions to predict the effect of average habitation size on the school attribute in question, and then combine this with the IV probit estimates of the effect of school attributes on enrollment to infer how habitation size affects enrollments through its effect on school inputs.

I graph the results separately for SC/ST children and other children, in figures 9 and 10 respectively. Each figure depicts two graphs. The first explores the effect of habitation size on enrollments, allowing habitation size to affect enrollment only through its effect on school access. The second graph adds on the additional influence of habitation size on school enrollment through its effect on teacher availability and school segregation.

Both graphs confirm the positive effect of habitation size on enrollment through school inputs. Children who live in large habitations are significantly more likely to be enrolled in school, not just because of the effect of size on school access, but also because of the positive effects of size on teacher strength.

Turning to the differences across castes, for scheduled castes and tribes, the additional effects of size on teacher availability and the availability of schools in SC/ST habitations imply a reduction in the effect of habitation size on enrollments. This primarily reflects the effects of school segregation. Districts with larger habitations which are characterized by improved access to schools are also characterized by a greater degree of segregation in schools. This reduces the schooling attainment, and hence the effectiveness of school location policies, for scheduled castes. The opposite is true for members of upper castes. For them, the advantage of living in districts with good access to schools is magnified considerably by the increase in school segregation which accompanies improved school access. Taken together, the results imply that districts in which habitation attributes imply generally better access to schools will also be characterized by greater intra-regional inequality in the schooling attainment of different castes.

8. Conclusion

In designing a policy for the construction and location of primary schools in rural India, the Government has paid scant attention to the fact that the placement of schools determines not just access, but also important dimensions of school quality. Because of the geographic distance between habitations in rural India and their small size, the decision to place a school within a habitation will determine total enrollment and hence the total number of teachers in the school. The policy of constructing schools in all habitations above a minimum size in conjunction with caste-based residential segregation also implies that many villages have multiple schools, with SC/ST children attending different schools from children of other castes.

Consistent with these hypotheses, this paper documents the significant role of habitation size in determining not just access to schools, but also critical school inputs such as the number of teachers. The inability to ensure that each school is of the optimal size, reflected in the effect of habitation size on teacher availability and the consequent variation in teacher strength across schools, implies that school location policies are partly responsible for the significant observed variation in school quality across rural India. Of habitations of size large enough to receive a school, households who reside in larger habitations will be characterized by higher levels of schooling than those who reside in smaller habitations.

Moreover, the results reveal that the average size of habitations in a district also dictates the availability of schools in SC communities, creating the potential for significant schooling segregation. Consistent with this hypothesis, I find that the effect of the availability of SC schools increases the schooling of other caste children, but reduces that of SC children. School location policies thus contribute to caste-based schooling inequality. The paper provides estimates of how the availability of schools in SC habitations changes the relationship between habitation size and schooling enrollment which would obtain if habitation size determined access to schools only. Without this effect, schooling attainment across castes would be far more equal.

By relating school location policies to school inputs associated with school quality and hence schooling attainment, the results of this paper suggest that improvements in school quality cannot be affected without re-considering school location policies. As rural India moves towards universal enrollment and schooling policy moves to address issues related to school quality, the relationship between school location policies and school quality need to be kept in mind.

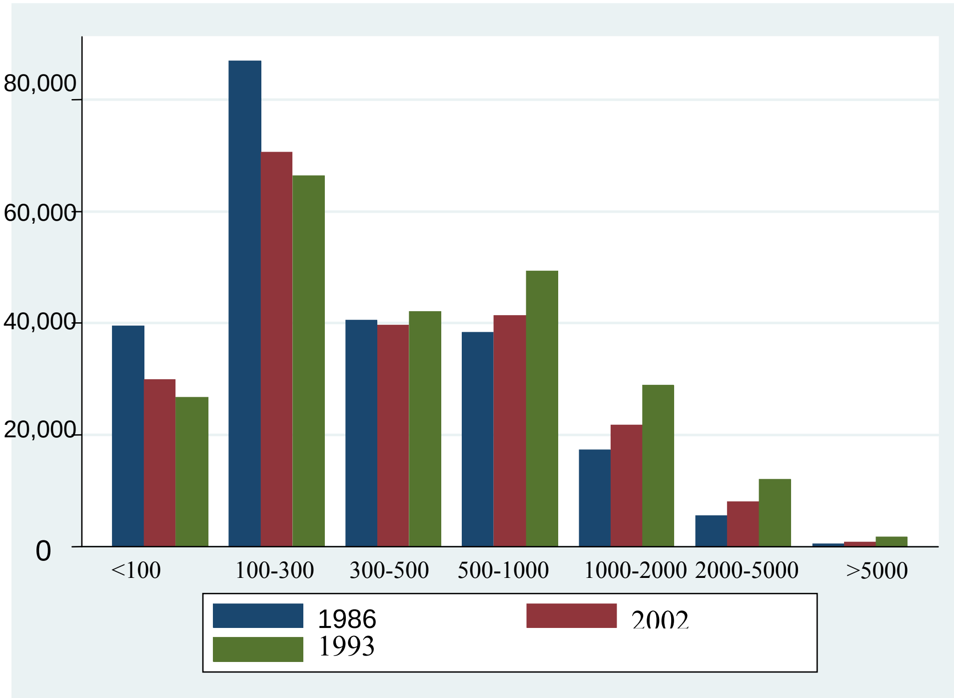

Figure 1: Distribution of habitations by habitation size, Uttar Pradesh, 1986-2002

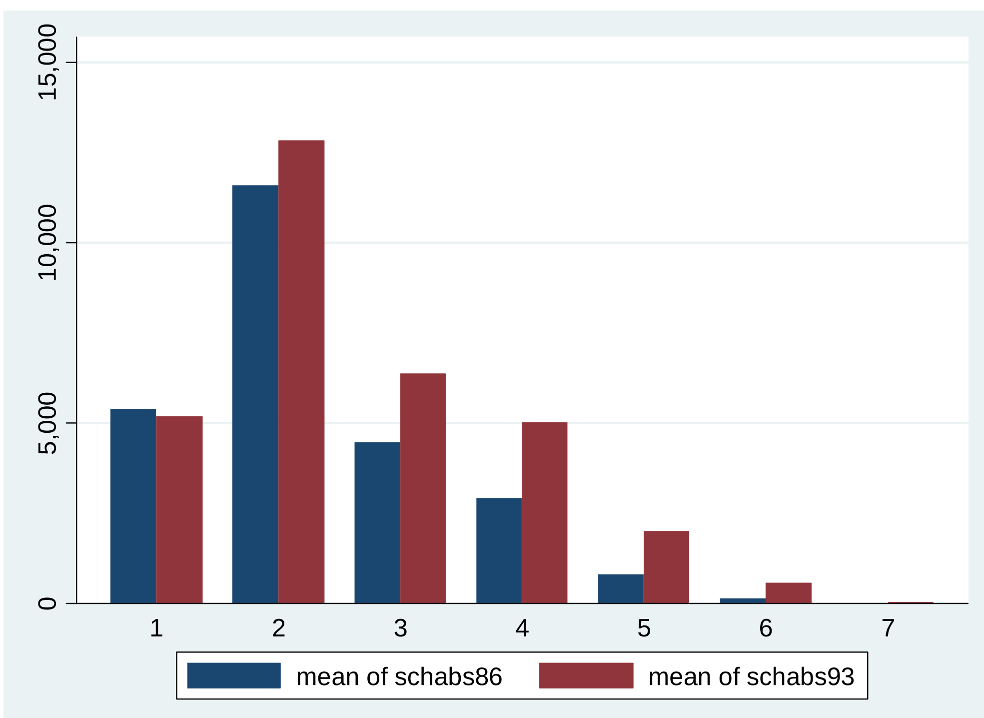

Figure 2: Distribution of SC and ST habitations by size, UP, 1986 and 1993

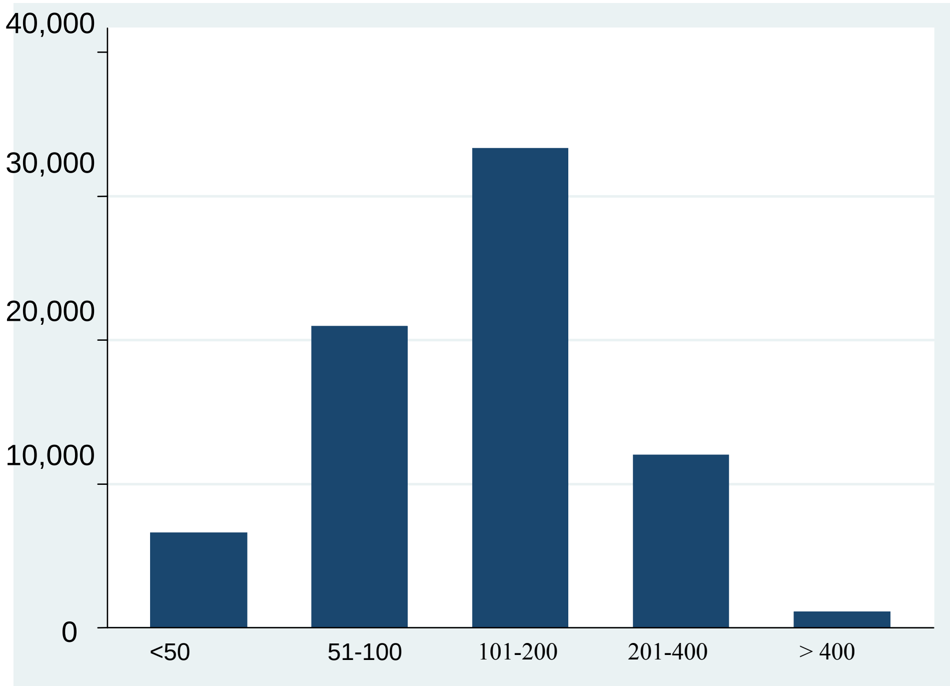

Figure 3: distribution of schools by school size, rural UP, 1993.

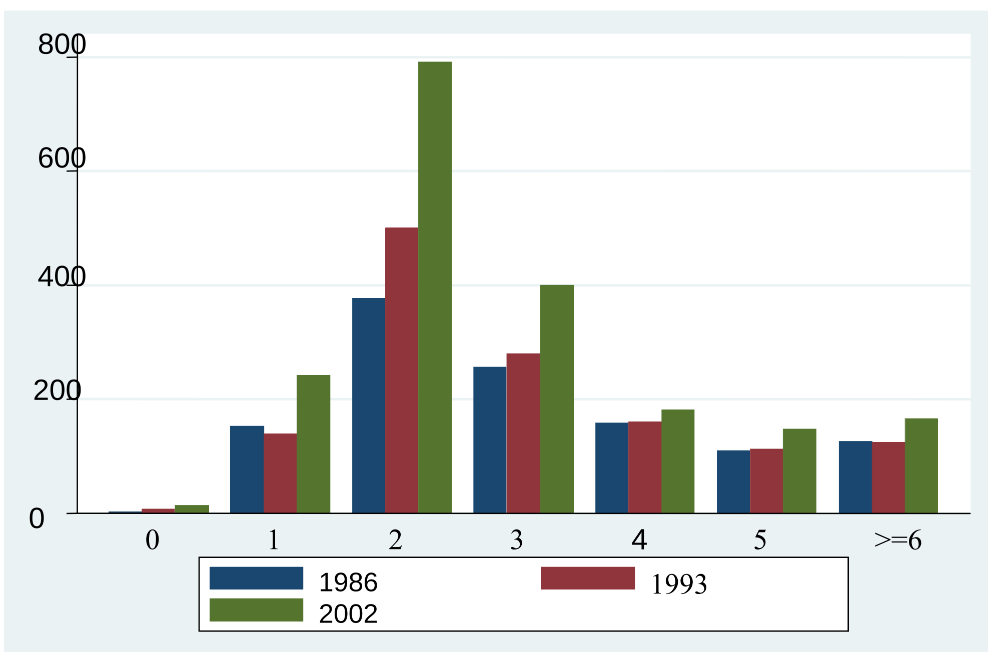

Figure 4: distribution of schools by number of teachers, rural UP, 1986-2002

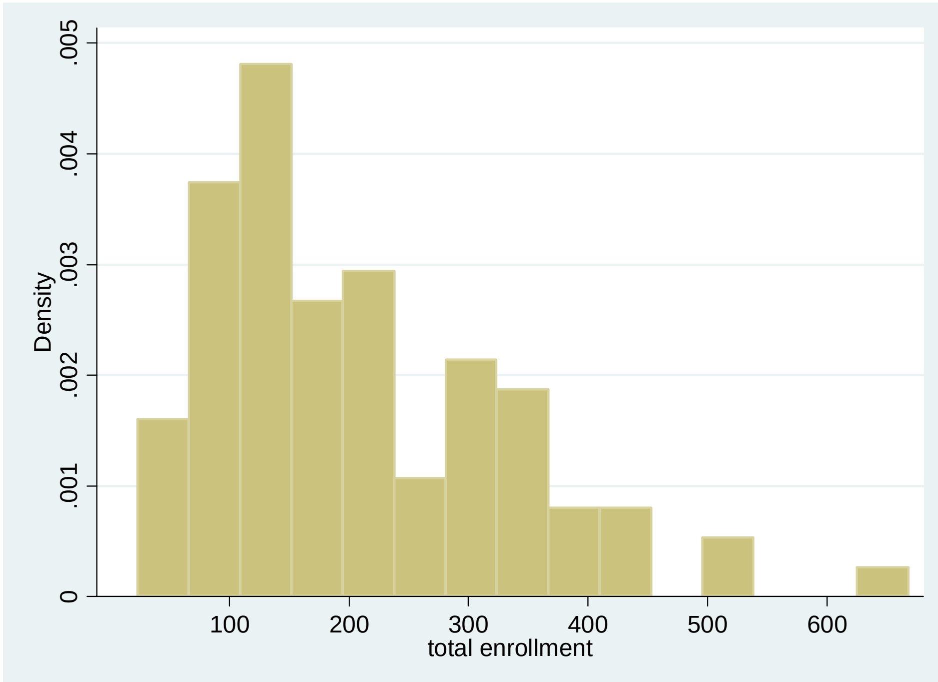

Figure 5: Distribution of schools by total enrollment, Ballia district, Block Bairiya, UP, 2002 (mean enrollment=206)

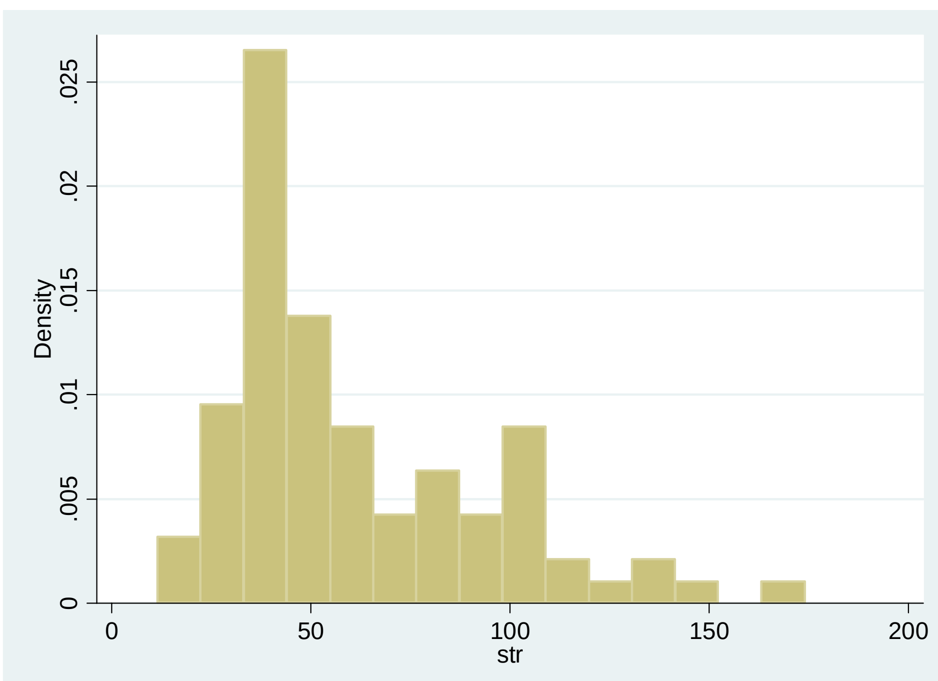

Figure 6: Student teacher ratios, District Ballia, Block Bairiya

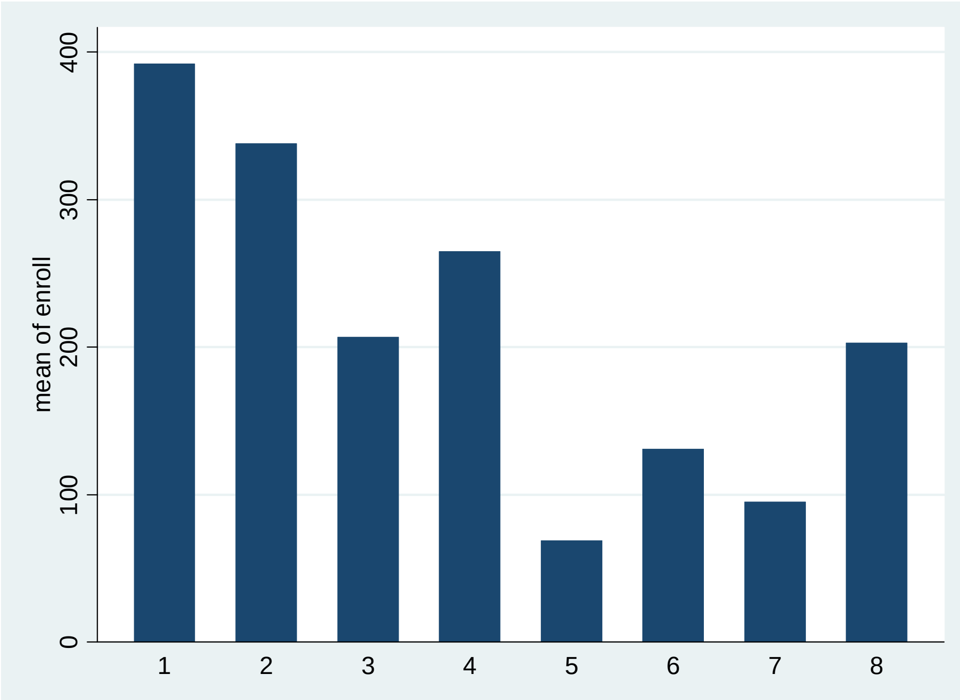

Figure 7: Distribution of schools by size, village …, District Ballia

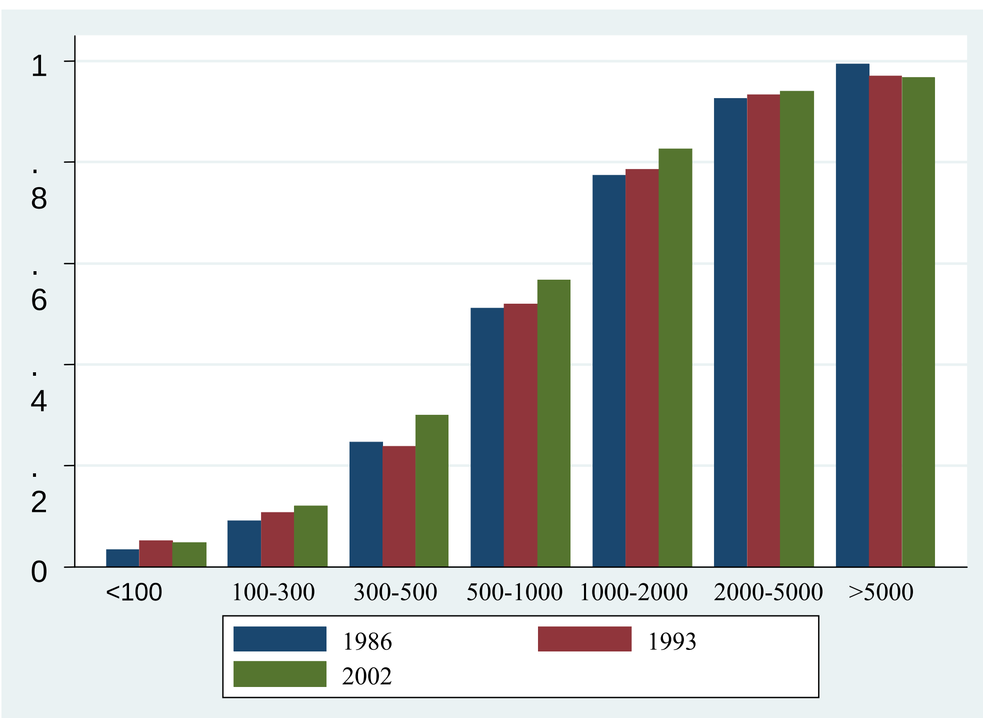

Figure 8: Proportion of habitations with schools in them, by habitation size, UP, 1986-2002

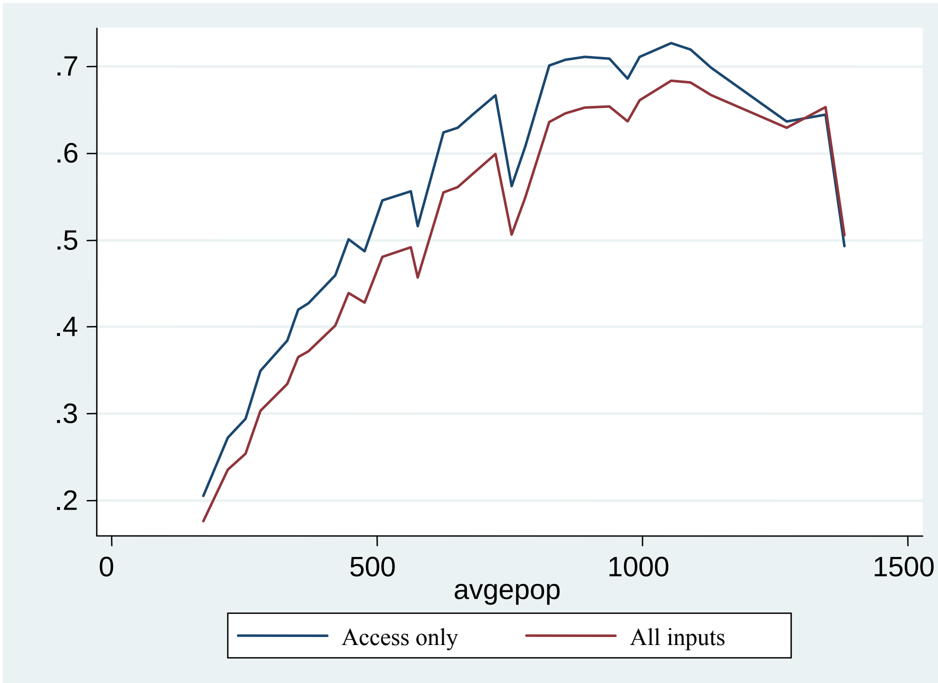

Figure 9: Effect of average habitation size on enrollment of SCST through school inputs

Figure 10: Effect of average habitation size on enrollment of other castes through school inputs

Table 1. Simple correlation coefficients between habitation size and school attributes

	sch_hab	Prop schools with <=2 teachers	Sch_SChab	Hab_5	Avge pop
Sch_hab	1.0				
Prop schools w/ <=2 teachers	0.25	1.0			
Sch_SChab	0.57	0.17	1.0		
Hab_5	0.73	-0.17	0.31	1.0	
Avge pop	0.70	-0.24	0.23	0.91	1.0

Note: Hab_5 is the proportion of habitations of population 500 or more. Sch_hab is the proportion of habitations with a school within the habitation, while Sch_SChab is the ratio of the number of SC/ST habitations with a school to total habitations. Avge pop is the average population size of a habitation.

Table 2. Simple regressions of School Attributes on habitation characteristics

	Pr sch_hab	Pr. Tchrs <=2	Pr Sch_SC hab
Pr hab_5	0.65*	0.86	0.008
	(0.25)	(0.59)	(0.09)
Pr SChab_5	0.43	-1.39	0.69*
	(0.63)	(1.37)	(0.29)
Pr hab_3	-0.79	0.54	-0.29+
	(0.28)	(1.01)	(0.16)
Pr SChab_3	-0.46	-1.69	2.14*
	(0.88)	(3.03)	(0.77)
Avge pop	-0.0007*	-0.001*	4.2 e-6
	(0.0003)	(0.0007)	(2.8 e-8)
Avge pop square	1.69 e-7*	2.96 e-7	-1.49 e-8
	(7.47 e-8)	(1.86 e-7)	(2.77 e-8)
# habitations	-0.0001*	-0.0002*	-0.00002*
	(0.00002)	(0.00007)	(7.2 e-6)
# SC habitations	-0.00002	0.0003	-0.00002
	(0.00007)	(0.0003)	(0.00002)
District p.c. exp	-0.0001	-0.0002	-0.00003
	(0.001)	(0.0003)	(0.00005)
District ed rank	-0.0001	-0.0005+	0.00006
	(0.0001)	(0.0003)	(0.00005)
Lagged # schools	0.0001*	0.001*	-1.32 e-6
	(0.0004)	(0.0001)	(0.00001)
Lagged enrollments	2.73 e-7	-4.6 e-6*	-5.79 e-9
	(2.74 e-7)	(9.1 e-7)	(1.08 e-7)
Dummy year 2	0.16*	-0.17+	0.04*
	(0.03)	(0.09)	(0.02)
Regression R^2	0.91	0.74	0.83
Sample size	110	110	110

Note: Standard errors are clustered at the district level.
* Significant at 5% level
+ Significant at 10% level

Table 3. Effect of school in habitation on enrollment in primary school, rural Uttar Pradesh, Children ages 6-12

	Proportion of habitations with schools	Currently enrolled in primary school		
		Probit	IV Probit	Marginal effect
Prop. Habitations with schools	--	0.83* (0.13)	2.28* (0.35)	0.48
Habitations 300-500*rank	-8.87 e-7* (2.14 e-7)	--	--	--
Prop. Habitations greater than 500* rank	-0.002* (0.0005)	--	--	--
Habitations 300-500	0.0001 (0.0001)	0.0001* (0.00006)	0.0003* (0.0001)	0.00006
Prop. Habitations greater than 500	0.89* (0.17)	-0.70* (0.18)	1.495* (0.256)	0.32
SC/ST	0.005 (0.003)	-0.03 (0.02)	-0.031 (0.025)	-0.007
Sex	-0.001 (0.001)	-0.34* (0.03)	-0.34* (0.03)	-0.072
Age	-0.004 (0.004)	1.43* (0.08)	1.43* (0.08)	0.302
Household p.c. exp	-2.57 e-6 (4.45 e-6)	0.0002* (0.00006)	0.0002* (0.0006)	0.00004
District avge p.c. exp	-0.0001 (0.0002)	-0.00007 (0.0002)	-0.0001 (0.0002)	-0.00002
Educational rank	0.001* (0.0004)	-0.0004* (0.0002)	-0.0005* (0.0002)	-0.0001
# habitations	-0.00005* (0.00002)	0.00002 (0.00002)	0.00004+ (0.00002)	0.00001
Avge pop of habitations	0.00002 (0.00004)	-0.00006 (0.00006)	-0.0001+ (0.00006)	-0.00002
District Prop. SC	0.026 (0.12)	-0.04 (0.19)	-0.16 (0.20)	-0.034
Sample size	16,759	16,759	16,759	
Test Statistic	F(29,61)=104.14	X^2=2004.48	χ^2=1843.64	
Wald test of exogeneity $X^2(1)$			20.35 (p=0.00)	

Table 4. Effect of school access by caste and gender
(Dependent variable: currently enrolled in primary school, ages 6-12)

	SC/ST	Other castes	Boys	Girls
Prop. Habitations with schools	3.39* (0.81)	1.94* (0.38)	1.81* (0.46)	2.86* (0.53)
Habitations 300-500	0.0002+ (0.0001)	0.0003* (0.0001)	0.0002+ (0.0001)	0.0005* (0.0001)
Prop. Habitations > 500	-2.41* (0.57)	-1.31* (0.29)	-1.20* (0.35)	-1.92* (0.38)
SC/ST	--	--	-0.05+ (0.03)	-0.01 (0.04)
Sex	-0.43* (0.05)	-0.32* (0.03)	--	--
Age	1.08* (0.16)	1.54* (0.09)	1.49* (0.11)	1.39* (0.12)
Hhold p.c. exp	0.0004* (0.0002)	0.0002* (0.0001)	0.0001 (0.0001)	0.0004* (0.0001)
District p.c. exp	-0.0003 (0.0004)	-0.0001 (0.0002)	0.0002 (0.0003)	-0.0004 (0.0003)
Education rank	-0.0002 (0.0004)	-0.0005* (0.0002)	-0.0005* (0.0003)	-0.0004* (0.0002)
# habitations	0.0001 (0.0001)	0.00005+ (0.00003)	0.0001* (0.00003)	0.00002 (0.00004)
Avge habitation population	-0.0001 (0.0001)	-0.0001+ (0.00007)	-0.0002+ (0.00009)	-0.00007 (0.0001)
Prop. SC	-0.05 (0.37)	-0.22 (0.23)	-0.18 (0.27)	-0.10 (0.29)
Sample size	3,863	12,896	9.143	7,616
Wald χ^2 (27)	425.82	1456.43	804.64	997.89
Wald test of exogeneity, $\chi^2(1)$	8.28 (p=0.004)	12.06 (p=0.001)	8.25 (p=0.004)	12.87 (p=0.0003)

Table 5. First Stage Regressions

	Proportion habitations with school	Proportion of schools with <=2 teachers	SC habitations with schools to total habitations
Habitations 300-499 * rank	-1.51 e-6*	-4.73 e-6*	-2.94 e-7*
	(4.43 e-7)	(1.10 e-6)	(1.26 e-7)
Prop. Habitations >500 * rank	-0.002*	-0.001	0.0003+
	(0.0008)	(0.002)	(0.0002)
Predicted prop of schools with <=2 teachers	-1.37 e-7	4.39 e-6*	-1.35 e-7
	(4.27 e-7)	(1.17 e-6)	(8.18 e-8)
Predicted <=2 teachers * rank	3.81 e-9	-1.56 e-8*	1.07 e-9
	(3.05 e-9)	(7.88 e-9)	(7.37 e-10)
Predicted prop <=2 teachers, square	-2.15 e-11	2.84 e-9*	-2.68 e-11
	(5.21 e-10)	(1.34 e-9)	(8.83 e-11)
Predicted prop <=2 teachers square* rank	2.09 e-12	-1.20 e-11	1.43 e-13
	(3.54 e-12)	(9.00 e-12)	(6.75 e-13)
SC habitations 300-500*rank	7.36 e-6	0.00001+	1.43 e-6
	(2.85 e-6)	(8.29 e-6)	(8.75 e-7)
SC habitations >500 * rank	-0.006	-0.04*	-0.003*
	(0.006)	(0.02)	(0.001)
Habitations 300-499	0.0001+	0.001*	0.00002
	(0.00006)	(0.0002)	(0.00002)
Prop habitations >500	0.292	-1.04+	-0.04
	(0.252)	(0.62)	(0.05)
SC habitations 300-499	-0.001*	-0.001	-0.0001*
	(0.0002)	(0.001)	(0.00007)
SC habitations >500	3.71*	5.16	1.50*
	(1.06)	(4.27)	(0.37)
# habitations	-0.0001*	-0.0001	-2.52 e-6
	(0.00002)	(0.0001)	(4.68 e-7)
Avge hab population	0.0002	0.001*	3.71 e-6
	(0.0001)	(0.0002)	(0.00003)
District prop. SC	-1.10*	2.90	0.06
	(0.53)	(2.47)	(0.19)
Sample size	16,759	16,759	16,759
Regression R^2	0.92	0.74	0.90

Table 6: Effect of School Access Controlling for Number of Teachers and School Segregation

	Probability of current enrollment in primary school		
	(1)	(2)	(3)
Prop. Habitations with schools	2.34*	3.59*	3.42*
	(0.39)	(0.42)	(0.44)
Prop. of schools with <=2 teachers	--	-0.35*	-0.23^{+}
		(0.07)	(0.13)
Prob. of school in SC habitation	--	4.59*	4.93*
		(1.58)	(1.61)
Habitations 300-500	-0.001*	0.0001	0.0001
	(0.0003)	(0.0001)	(0.0001)
Prop. Habitations greater than 500	-3.05*	-1.31*	-0.97*
	(1.61)	(0.26)	(0.39)
SC habitations 300-500	0.001*	0.001*	0.001*
	(0.0003)	(0.0003)	(0.0003)
Prop. SC habitations >500	-3.05^{+}	-11.02*	-10.43*
	(1.61)	(2.23)	2.38
SC/ST	-0.03	-0.06*	-0.06*
	(0.02)	(0.03)	(0.03)
Sex	-0.34*	-0.34*	-0.34*
	(0.03)	(0.03)	(0.03)
Age	1.4 4*	1.45*	1.45*
	(0.08)	(0.08)	(0.08)
Household p.c. exp	0.0002*	0.0002*	0.0002*
	(0.00006)	(0.0001)	(0.0001)
District avge p.c. exp	0.0002	0.0004*	0.0005*
	(0.0002)	(0.0002)	(0.0002)
Educational rank	-0.0003	-0.0009*	-0.0008*
	(0.0002)	(0.0002)	(0.0002)
# habitations	0.0001*	0.0002*	0.0001*
	(0.00003)	(0.0003)	(0.00003)
# SC habitations	-0.0002*	-0.001*	-0.0005*
	(0.0001)	(0.0001)	(0.0001)
Avge pop of habitations	-0.0002^{+}	-0.0003*	-0.0005*
	(0.0001)	(0.0001)	(0.0002)
Avge pop SC habitations	-0.0002	-0.0003^{+}	-0.0003^{+}
	(0.0001)	(0.0002)	(0.0002)
District Prop. SC	-1.08	0.35	0.03
	(0.95)	(1.02)	(1.06)
Predicted prop teachers<=2	--	--	-3.31 e-7
			(4.07 e-7)
Predicted prop teachers <=2 squared	--	--	-9.34 e-11
			(2.62 e-10)
Sample size	16,759	16,759	16,759
Wald χ^2	1871.72	1883.91	1884.43
Wald test of exogeneity X^2	17.57	91.02	87.36

Table 7.: Robustness Check: Including square of avge population of habitation

	First stage regressions			Current enrollment
	Prob. school in habitation	Prop. Teachers <=2	Prob SC hab with school	
Prob. School in habitation	--	--	--	3.87*
				(0.44)
Prop. Teachers <=2	--	--	--	-0.32*
				(0.06)
Prob. School in SC habitation	--	--	--	2.77*
				(1.44)
Habitations 300-499 * rank	-1.39 e-6*	-3.74 e-6*	-3.31 e-7*	--
	(4.18 e-7))	(9.60 e-7)	(1.24 e-7)	
Prop. Habitations >500 * rank	-0.002**	-0.0001	0.0003+	--
	(0.0007)	(0.002)	(0.00018)	
Predicted prop of schools with <=2 teachers	-2.64 e-7	3.3 e-6*	-8.37 e-8	--
	(4.65 e-7)	(9.56 e-7)	(8.99 e-8)	
Predicted <=2 teachers * rank	4.54 e-9	-9.33 e-9	7.89 e-10	--
	(3.28 e-9)	(6.57 e-9)	(7.41 e-10)	
Predicted prop <=2 teachers, square	-3.3 e-10	1.82 e-10	9.27 e-11	--
	(6.64 e-10)	(1.08 e-9)	(1.12 e-10)	
Predicted prop <=2 teachers square* rank	3.60 e-12	1.02 e-12	-4.24 e-13	--
	(4.23 e-12)	(7.14 e-12)	(7.54 e-13)	
SC habitations 300-500*rank	7.19 e-6*	0.00001*	1.46 e-6+	--
	(2.92 e-6)	(6.02 e-6)	(8.18 e-7)	
SC habitations >500 * rank	-0.005	-0.03	-0.003*	--
	(0.005)	(0.02)	(0.001)	
Habitations 300-499	0.0001	0.0004	0.00002	0.0002
	(0.0001)	(0.0002)	(0.00002)	(0.0001)
Prop habitations >500	0.48*	0.54	-0.11*	-1.17*
	(0.23)	(0.42)	(0.05)	(0.41)
SC habitations 300-499	-0.001*	-0.0007	-0.0001+	0.001*
	(0.0003)	(0.0006)	(0.00007)	(0.0003)
SC habitations >500	3.70*	5.00	1.49*	-8.00*
	(1.06)	(3.96)	(0.34)	(2.12)
# habitations	-0.00005*	-0.0001	-3.27 e-6	0.0002*
	(0.00002)	(0.0001)	(4.6 e-6)	(0.00003)
# SC habitations	0.0002*	0.0002	-7.08 e-6	-0.0005*
	(0.00005)	(0.0001)	(0.00001)	(0.0001)
Avge hab population	-0.0002	-0.003*	0.0002*	-0.001*
	(0.0004)	(0.0007)	(0.0001)	(0.0005)
Avge hab population square	1.40 e-7	1.21 e-6*	-6.15 e-8*	3.75 e-7*
	(1.24 e-7)	(2.42 e-7)	(2.95 e-8)	(1.42 e-7)
Avge pop SC habitation	-0.00001	-0.0001	-0.0001	0.001*
	(0.0003)	(0.001)	(0.0001)	(0.0005)
Avge SC pop square	-6.62 e-10	-2.64 e-8	2.39 e-8	7.84 e-7*
	(1.68 e-7)	(4.11 e-7)	(3.07 e-8)	(2.35 e-7)
Sample size	16,759	16,759	16,759	16,759
Regression R^2	0.92	0.80	0.90	Wald χ^2=1893.05

Table 8: Enrollment regressions by caste

	SC/ST	Other castes
Prop. Habitations with schools	4.30* (0.87)	3.24* (0.48)
Prop. of schools with <=2 teachers	-0.21+ (0.13)	-0.35* (0.08)
Prob. of school in SC habitation	-2.34 (3.00)	6.45* (1.90)
Habitations 300-500	-0.0002 (0.0002)	0.0002 (0.0001)
Prop. Habitations greater than 500	-1.59* (0.51)	-1.25* (0.31)
SC habitations 300-500	0.0005 (0.0006)	0.001* (0.0003)
Prop. SC habitations >500	-4.48 (4.77)	-12.67* (2.70)
Sex	-0.43* (0.05)	-0.32* (0.03)
Age	1.09* (0.07)	1.55* (0.09)
Household p.c. exp	0.0005* (0.0002)	0.0002* (0.0001)
District avge p.c. exp	0.0005 (0.0004)	0.0003+ (0.0002)
Educational rank	-0.0002 (0.0005)	-0.0009* (0.0003)
# habitations	0.0002* (0.0001)	0.0001* (0.00004)
# SC habitations	-0.0004+ (0.0002)	-0.0005* (0.0001)
Avge pop of habitations	-0.00004 (0.0002)	-0.0004* (0.0001)
Avge pop SC habitations	-0.001* (0.0003)	-0.0001 (0.0002)
District Prop. SC	2.91 (2.07)	-1.23* (0.31)
Sample size	3863	12896
Wald χ^2	448.71	1480.90
Wald test of exogeneity χ^2	18.21 (0.0004)	70.87 (0.0000)

References

Alderman, Harold, Jere R. Behrman, Victor Lavy, Rekha Menon. 2001. "Child health and social enrollment: A longitudinal Analysis." *The Journal of Human Resources* 36(1): 185-205.

Benabou, R. 1996. "Equity and Efficiency in Human Capital Investment: The Local Connection." *Review of Economic Studies* 63:237-264.

Birdsall, N. 1985. "Public Inputs and Child Rearing in Brazil." *Journal of Development Economics* 18(1):67-86.

Borjas, George J. 1995. "Ethnicity, Neighborhoods, and Human-Capital Externalities." *American Economic Review* 85(3): 365-390.

Case, A. and Katz, L. 1991. "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youth. NBER Working Paper 3705.

Coleman, J. 1988. "Social Capital in the Formation of Human Capital." *American Journal of Sociology* 94:S95-S120.

Deolalikar, A. 1997. "Determinants of School Enrollment and School Expenditures in Kenya." Manuscript.

Filmer, D. 2004. "If You Build It, Will They Come? School Availability and School Enrollment in 21 Poor countries." World Bank Policy Research Working Paper 3340. Washington, D.C. The World Bank.

Fernandez, Raquel and Richard Rogerson. 1996. "Income Distribution, Communities, and the Quality of Public Education." *Quarterly Journal of Economics* 111(1): 135-164.

Foster, Andrew D. and Mark R. Rosenzweig. 1996. "Technical Change and Human Capital returns and Investments: Evidence from the Green Revolution." *American Economic Review* 86(4):931-53.

Kochar, Anjini, Kesar Singh and Sukhwinder Singh. 2007. "Targeting Public Goods to the Poor in a Segregated Economy: An Empirical Analysis of Central Mandates in Rural India." Manuscript.

Miguel, Edward and Mary Kay Gugerty. 2995. "Ethnic Diversity, Social Sanctions and Public Goods in Kenya." *Journal of Public Economics* 89(11-12):2325-2368.

Lavy, Victor. 1996. "School Supply Constraints and Children's Educational Outcomes in Rural Ghana." *Journal of Development Economics* 51:291-314.

Lillard, L. and R. Willis. 1994. "Intergenerational Educational Mobility: Effects of Family and State in Malaysia." *Journal of Human Resources* 29(4):1126-1166.

Oates, W. and Schwab, R. 1991. "Community Composition and the Provision of Local Public Goods: A Normative Analysis." *Journal of public Economics* 44:217-238.