



Stanford

Center for International
Development

Working Paper No. 354

School Size and Schooling Inequalities

by

Vergheze Jacob

Anjini Kochar

Suresh Reddy

November 2008



Stanford University
John A. and Cynthia Fry Gunn Building
366 Galvez Street | Stanford, CA | 94305-6015

SCHOOL SIZE AND SCHOOLING INEQUALITIES

Verghese Jacob,[†] Anjini Kochar^{††} and Suresh Reddy[†]

November 2008

Abstract

Despite the very small school sizes observed in developing economies, including India, there has been almost no attention paid in the economics literature to its potential effect on learning or on schooling inequalities. Small school size, and the consequent reduction in the number of teachers, requires schools to combine grades in multi-grade classrooms. Because of the residential segregation by caste which characterizes rural India, children of scheduled castes and tribes, who generally live in smaller, segregated habitations than their upper caste counterparts, are more likely to experience multi-grade teaching. Combining panel data with an instrumental variable strategy which enables us to control for cohort- and school specific determinants of quality which may otherwise bias estimates of the effect of classroom attributes, we find that multi-grade teaching significantly reduces schooling achievement and contributes to caste based schooling inequalities. Our methodology also allows us to obtain estimates of the effect of classroom size and caste composition which are free of traditional sources of bias, and to compare these to those of multi-grade teaching.

Keywords: India, education, schooling inequalities

JEL Classification No.: I21, C33

[†]Byrraju Foundation, India

^{††} Corresponding author, Stanford University

This paper uses data collected by the Byrraju Foundation. The authors wish to thank the many researchers at Byrraju Foundation who worked tirelessly on this project.

1. Introduction

The question of what explains the persistent differences in schooling across households distinguished by race or ethnicity has long absorbed the attention of researchers. In India, these differences take the form of differences by caste. Children of “scheduled” castes and tribes (SC/STs), which comprise the lowest rung of the caste hierarchy, have consistently lower levels of schooling than members of “upper” castes. The most recent household data collected by the Government of India (2004), for example, reveal that in rural areas of the southern state of Andhra Pradesh, our study area, the proportion of 12 to 18 year olds with 5 years of schooling is 0.78 for children from upper castes, but only 0.68 for those from scheduled castes and tribes.¹ Suggestive of a higher drop out rate on completion of primary schooling for members of scheduled castes, the gap widens at higher schooling levels; the percentage of children aged 15-18 with at least 8 years of schooling being 63% for upper castes, in comparison to 49% from scheduled castes and tribes.

What explains this gap? As in the case of Black-White schooling differentials in the United States, one would suspect that differences in the quality of schools attended by members of scheduled castes, relative to other castes, has much to do with it. Mirroring conditions in the U.S. economy, rural India is characterized by extensive residential segregation, with households belonging to scheduled castes and tribes residing in separate sub-habitations of the main village. In order to ensure access to schools, particularly for the poor, the Government of India’s school location policy takes the habitation, rather than the village, as the unit of planning, and aims to provide a school in each habitation, provided it

¹ Government of India, National Sample Survey Organization, 61st round.

has a population in excess of 300. For scheduled caste habitations, this population threshold is only 200. As a consequence of this policy, 78% of habitations in India, and as many as 96% in Andhra Pradesh, have a school located within the habitation. Residential segregation by caste, in conjunction with the policy of providing a school to each habitation, implies that scheduled caste households generally attend separate schools from members of upper caste. It is natural, then, to initiate any analysis of caste differences in schooling with a consideration of the role of differences in school quality.

This research builds on the voluminous research on the Black-White schooling gap in the United States. Fryer and Levitt (2004, 2005), Hanushek and Rivkin (2006), and others, document the significant role of differences in school quality in explaining this differential. The specific school qualities that have been the focus of research are the “traditional” factors known to explain schooling achievement, such as class-size, the ethnic composition of students, and teacher qualities such as experience and education. Hanushek and Rivkin, for example, find that differences across schools in characteristics such as student racial composition, teacher experience and student turnover account for all the growth in the achievement gap following third grade.

In contrast, there has been hardly any economic research which examines the determinants of ethnic differences in schooling in developing economies. While the case for school quality explanations is compelling, qualities such as class size and teacher attributes may play less of a role than they do in the United States. In India, for example, class size is determined by an allocation rule which seeks to provide one teacher per 40 students. This does not guarantee equality in class size – even if the target ratio was not

exceeded in any school, class size would still vary, and the variation would be greater in schools with lower total enrollment. However, schooling policies in India also call for a minimum of two teachers per school which means that, if anything, average class sizes are smaller in the smaller-than-average habitations populated by scheduled castes and tribes. Differences in observable teacher attributes may also only minimally affect schooling inequality. Unlike the United States, decisions regarding teacher hiring, appointment to schools, etc. are not determined by local institutions, but by higher level district governments. Most state governments follow a teacher transfer policy whereby teachers are transferred across schools in the district every three years. Because of this policy, there is little significant variation in teacher quality across schools as measured by observable factors such as experience.

A far more compelling source of differences in school quality is differences in *school* size, as measured by total enrollment. Because of the government's decision to provide schools in even relatively small habitations, average school sizes are small. In 2002, the average size of a school in rural India was 244, while that of a school in the state of Andhra Pradesh was 196. And, because scheduled caste and tribe habitations are, on average, significantly smaller than upper caste habitations, one of the more striking differences in schools across these two types of habitations is the difference in the size of the average school.

Small school sizes imply an inability to exploit economies or scale, and hence reduced investment in infrastructure such as science labs and playground equipment. However, perhaps the most detrimental consequence of small school size is the reduction

in the number of teachers. Because teacher assignments are based on enrollment, small school sizes imply that the majority of schools have less than 5 teachers, the minimum required to ensure one teacher per grade. Indeed, 64% of schools in rural India, and 66% in Andhra Pradesh, have two or fewer teachers. Consequently, “multi-grade” teaching, which combines the students of one grade with that of another in a single classroom, is common. And, it is particularly prevalent in scheduled caste and tribe habitations which are, on average, significantly smaller in size than upper caste habitations.

Many schools in developed countries favor a multi-grade environment, particularly in lower grades. However, when it is instituted in response to low total enrollments, the decision on which grades to combine in one classroom is generally made on the basis of grade-level enrollments, resulting in combinations which are not optimally selected from the viewpoint of learning strategies. In the sample of schools on which this study is based, it is not uncommon to see first grade students being jointly taught with students in the 3rd, 4th and 5th grades.

Even in this context, multi-grade teaching need not be detrimental to learning if teaching followed a flexible, experimental and group-based approach. Indian government schools, however, follow a textbook, curriculum-based approach, with a clear and distinct curriculum for each grade. As a consequence, teaching in a multi-grade environment implies a substantial reduction in the time the teacher devotes to the teaching of the curriculum for any one grade. In a classroom in which first grade students are combined with those of fifth grade, the teacher essentially devotes half her time to the teaching of the

first grade students and the other half of her time to the teaching of fifth grade students. As a consequence, one would expect learning to be significantly reduced.

Despite the very small average school sizes in India (and many other developing economies), there has been almost no attention paid to its adverse effects on learning, or its potential for generating schooling inequalities. This perhaps reflects the fact that much of the existing literature on ethnic or racial schooling differentials examines the case in the U.S. economy. In the U.S., as in most developed economies and even the urban areas of developing economies, school sizes are such that most grades are further divided into classrooms – multi-grade teaching is not resorted to, unless it is optimal from the viewpoint of the curriculum and learning strategies adopted in the school in question. Consequently, the available literature has been more occupied with identifying the effect of attributes such as class size and the racial composition of students, ignoring the potential contribution of school size. In this paper, we argue that in economies such as India, school size is likely to be as important a factor in explaining schooling outcomes as are traditional factors such as the racial or ethnic composition of the student body. This is also likely to be true of other developing economies characterized by relatively small rural habitations and, consequently, schools.

Research on the effect of classroom attributes on learning must deal with the selective placement of children in schools and classrooms; as a consequence of selection, variation in caste composition or classroom size across schools may reflect underlying variation in the demand for schooling across communities and households. Recent research on racial schooling inequalities uses data on multiple grades within a school to implement

fixed effect models, exploiting variation in the variable of interest within a school. A further improvement is possible with panel data, which enables the inclusion of grade-by-school fixed effects in regression equations and hence controls for (time-invariant) grade-level variation in quality.

Our study is based on data we collected for a sample of 60 schools in rural Andhra Pradesh. The survey provides information on multiple grades within a school over a two year period, allowing us to mimic existing research strategies. Because schooling conditions in India, fully described in the body of this paper, reduce the variation in grade-specific schooling quality across schools, we concentrate on cohort-level variation in quality within a school, using cohort-by-school fixed effects to identify the effect of multi-grade teaching by the change in the classroom environment experienced by any specific cohort in a school over the two years of the survey.

We go further, however, and also report results from an instrumental variables (IV) strategy, which recognizes that the probability of a child being placed in a multi-grade classroom reflects the joint probability of the school-level decision to have at least one multi-grade class, and the grade-specific decision on which grades to combine. Because only schools with less than 5 teachers must resort to multi-grade teaching, a specific enrollment level determines the school-level decision: multi-grade classrooms are never observed in schools with enrollments over the threshold size which ensures at least 5 teachers. At the grade level, we build on the fact that the decision on which grades within a school to combine in a classroom is based, amongst other factors, on the relative size rank of each grade. In an approach which parallels a conventional difference-in-difference

estimator, the identification of multi-grade classrooms then exploits the variation in cohort size within schools, recognizing that this variation determines multi-grade teaching only in schools whose enrollments fall below the cut-off level which determines the necessity for multi-grade classrooms in any school.

Multi-grade teaching provides an additional source of variation in classroom size and in the caste composition of students in a classroom, the two factors which have received the most attention in the research on supply side determinants of schooling. Recognizing this, we contribute to this literature, providing estimates of the effect of class size and caste composition which exploit only the variation in these variables which results from combining different grades in one classroom. While methodological details are deferred to the body of this paper, our research provides credible estimates of the effect of these attributes on learning for a developing economy, and contrasts them to those of multi-grade teaching.

To preview our results, we find that multi-grade teaching and the caste composition of students are equally important in explaining caste-based schooling inequalities in mathematics. For the teaching of language, however, it is multi-grade instruction which adversely affects schooling inequality. While large classroom size reduces learning by more than either multi-grade classrooms or the caste composition of the student body, there is little variation in reported class size by caste. Consequently, classroom size cannot explain caste-based schooling inequalities.

The remainder of this paper is structured as follows. Section 2 describes the survey data we collected for the purposes of this study. Conventional approaches to identification

are discussed in Section 3, while Section 4 specifies the empirical approach of this paper. The main empirical results of the paper are in Section 5, and their implications for caste-based schooling inequalities are considered in Section 6. The last section concludes.

2. Data

This paper draws on a survey of 60 schools the authors conducted in 2004 and 2005. The schools are located in the districts of East and West Godavri, in the Southern state of Andhra Pradesh. The survey region is a relatively prosperous part of the state, itself amongst the fast growing regions of India. The sample of 60 schools is divided into 45 “main village” schools and 15 hamlet schools. The data were collected by a team of researchers from the Byrraju Foundation, a non-governmental organization which operates in over 140 villages in this region. Surveyed villages were randomly chosen from the set of Program villages and, of the survey villages, 15 were randomly selected for inclusion in the “hamlet” sample.

In each village, we collected data on the socio-economic characteristics of the habitations or hamlets that comprised the village. In addition, we collected school-level data on government primary schools in the main village and in those hamlets selected for the survey. The school data provide information on enrollments (for the years 2003-04 to 2005-06) by grade and by caste, on teachers and teacher characteristics, and on schooling resources. Information on the grade or grades taught by all teachers allows us to match teachers to grades and, specifically, to identify multi-grade classes. For each multi-grade classroom, we know the specific grades which were combined in a single classroom.

An additional module collected results from language (Telugu) and mathematics tests for children in grades 2 through 4 in 2004-05, and grades 2 through 5 in 2005-06. For 2004-05, we have initial test scores from tests conducted at the start of the school year (August), a half-yearly scored (December), and scores for final examinations in April. For 2005-06, we have results for the August and December exams, but not for the final April exam.

As is the case in many developing economies, India does not support a system of standardized tests at each grade level. However, local authorities are required to regularly test students in language and mathematics. Though the tests are not standardized at the national level, there is nevertheless some level of uniformity since village schools are overseen by a three tier state-level organization, of which the state educational department constitutes the first tier, the district administration the second, and the block education office² the third. Examination papers for school-level exams are set at the block level, but overseen by the district government to ensure uniformity in standards. However, the results are graded by school teachers, leaving open the possibility that any particular school may inflate its results. While the block level government oversees and monitors exams to minimize such an event, schools located in scheduled caste hamlets regularly score well on these tests, relative to their counterparts in the main village, raising concerns that there is a strong school-specific component to test results. To allow for this, our analysis is based on variation in test scores within a school, with school fixed effects included in all regressions.

² The “block” government in Andhra Pradesh is referred to as the “Mandal.”

Before discussing the survey data, we provide information on habitation size across scheduled caste and upper caste habitations, for India, the survey region (the districts of West and East Godavri), and the survey villages (table 1). Data for India and the survey region are from the 7th All India Educational Survey (2002). In India, as in the survey region, each village comprises an average of two habitations, though the number of habitations in the survey village, at 4, is higher. The data in table 1 reveal the significant difference in the size of scheduled caste and upper caste habitations. For India as a whole, the average population per scheduled caste habitation is 425 versus 758 in upper caste habitations. Villages in our survey are substantially larger than the national average, but even here these differences in population size persist: The average population size of a scheduled caste habitation, at 1,120 is less than one-third that of an upper caste habitation (3,696). The data also provide some information on the size distribution of habitations, revealing the very large number of small habitations in the country. In India, 62% of all habitations have a population of less than 500. For scheduled caste habitations, this percentage is 74%.

Because habitations are relatively small, the Government's policy of providing a primary school to each habitation has resulted in schools in rural India also being relatively small. This is strikingly revealed in figure 1, which graphs the size distribution of *all* rural schools in the 14 blocks of West Godavri district where our survey was conducted. Of the total 619 schools in these blocks, the mean enrollment was 102, with the median size being only 80.

Table 2 provides descriptive statistics on government primary schools included in our survey, separately for schools located in the main village and those in scheduled caste hamlets (the latter are referred to as hamlet schools). The data reveal the considerable differences in the demographic and socio-economic profiles of students in these two types of schools, reflecting corresponding differences in the profile of their “catchment area,” generally the habitation in which they are located.³ First, the difference in the population size of habitations gets translated into differences in total enrollment. The mean size of the student body in main village schools was 154, as compared to 83 in hamlet schools. The assignment of teachers to schools is based on total enrollments, so the difference in school size is reflected in a corresponding difference in the number of teachers; on average, main village schools have 5 teachers, whereas hamlet schools have only 3. This implies that multi-grade teaching is the norm in hamlet schools. The data reveal that 82% of hamlet schools report at least one class featuring multi-grade instruction. This percentage is much lower (39%) in main-village schools.

A second difference between schools in the main village and those in hamlets relates to differences in the caste composition of their student population. The considerable residential segregation by caste across habitations is reflected in extensive schooling segregation. In hamlet schools, as many as 68% of the student body are members of scheduled castes and tribes. In contrast, only 17% of students in main village schools come

³ Because over --% of habitations have a school located within the habitation, the school catchment area is generally synonymous with the habitation in which it is located. Children are, in principle, able to attend any government school they choose. However, transferring out of the catchment area requires an application and the completion of an approval process at the block level. Because of the time costs and bureaucratic delays associated with this, it is rare to find students attending government schools in habitations other than their own – there are no such cases in our survey.

from these castes. Upper caste children who attend main village schools are surrounded primarily by other upper caste students.

Turning to teacher characteristics, table 2 reveals that while the mean age of teachers, and hence their years of teaching experience, does not vary significantly across the two types of schools, there are considerable differences as regards other socio-economic characteristics. Teachers in hamlet schools are more likely to be members of scheduled castes (31% as compared to 21% in main village schools). They are also more likely to be male (60% as compared to 53%). Finally, they have significantly lower educational levels – while 83% of teachers in main village schools have either a B.A. or a M.A. degree, only 69% of the teachers in hamlet schools do.

Table 3 provides more information on multi-grade teaching. For each grade, the table provides the probability of being combined with other grades. Thus, 36% of grade 1 classes in our sample are jointly taught with grade 2 students. The data reveal that while adjoining grades are most commonly combined in one multi-grade classroom, it is not unusual to find first and second grade students combined with those from fourth and fifth grade. For example, 26% of first grades are jointly taught with students from grades 3, 4 or 5. Similarly, 18% of grade 5 students are combined in a classroom with students from the first to the third grade. The reason is that the choice of which classes to combine appears to be strongly driven by the motivation to equalize class sizes across teachers. As a consequence, smaller classrooms are far more likely to be combined in multi-grade environments, even if this results in non-optimal combinations from a teaching perspective.

3. Conventional Approaches to Identification

Empirical analysis of the effects of classroom attributes, such as class size or the racial or ethnic composition of students in a class, on individual achievement is rendered difficult by the fact that the placement of students in classrooms and schools is not random, but reflects observed and unobserved attributes of the student, his or her family and cohort, teachers, and the school itself. If these attributes directly affect learning, then simple OLS estimates of the effect of, for example, the racial composition of the class on individual learning will be biased. The cross-sectional variation in the proportion of scheduled caste students in a classroom, in data which combines observations on classrooms in different schools, could reflect underlying characteristics of the community which affect schooling. Or, classrooms with a higher ratio of students from disadvantaged races may be assigned better quality (or poorer quality) teachers. In either case, it is difficult to separate out the pure effects of the caste composition of students from that of other unobserved determinants of schooling.

The effects of unobserved time-invariant attributes of the school or the community can be simply accommodated, if information is available for multiple grades within a school. This allows the inclusion of school fixed effects in the regression equation, with regression results reflecting the variation in classroom attributes across grades within a school (Lavy and Schlosser 2007).

While this removes the correlation of the classroom attributes in question with unobserved school-level measures of quality, it does not allow for grade- or cohort-specific variation in quality within a school. Both sources of variation are likely to exist in the data.

More attention may be paid to the development of the curriculum in higher grades, resulting in improved school quality at these levels. Or, the assignment of teachers could respond to observed or unobserved attributes of a cohort. Larger cohorts may be assigned better teachers. If so, regressions of class size on individual attainment will combine the pure effects of class size with those of (unobserved) teacher quality. The same result obtains if teachers are assigned to cohorts or grades on the basis of attributes other than size, such as the mean ability of the cohort, but if the quality of teachers influences class size through its effect on drop out rates or transfers to other schools. Here, the endogenous variation in class size will bias regression estimates.

The recent literature on the topic has primarily concentrated on bias caused by grade-specific variation in school quality within a school. Using grade-by-school fixed effects to wipe out the influence of any time invariant quality differences across grades within a school, Hanushek and Rivkin (2006) and Hanushek, Kain and Rivkin (2002), for example, identify the effects of the racial composition of students in the United States on schooling achievement by following the progress of different cohorts of students in a given school as they pass through the same grade. In their research, they relate variation in the achievement levels of different cohorts of students in a particular grade (of a specific school) to cohort variation in the black-white student ratio, augmenting the regression to include controls for cohort differences in movement across schools.

Research which purges individual achievement of the effect of grade- and school-level differences in quality, relying only on variation across adjacent cohorts in the attributes of concern, must still address biases introduced by cohort-specific variation in

quality due to either time-invariant or time-varying factors. School teachers, for example, commonly describe a particular cohort as being brighter than others, or as more disciplined. Just as there may be random variation in the size of adjacent cohorts, there may also be random variation in ability or in other character traits which affect schooling achievement. A more gifted cohort may experience lower drop out rates and may also benefit from having better quality teachers assigned to it. This suggests that any correlation between cohort attributes and student achievement may reflect the effect of unobserved variation in mean cohort ability.

While using cohort-by-school fixed effects could similarly purge regression estimates of the effect of unobserved time-invariant cohort traits, this methodology has rarely been used to study the effects of classroom traits such as size or student composition. The inclusion of cohort-by-school fixed effects implies that the effects of the racial composition of the class, for example, is identified only as a consequence of variation in this composition for a particular cohort over time. In most instances, such changes can only be an endogenous response to school or grade quality, or other time-varying determinants of schooling, such as schooling returns for the cohort in question. Dealing with this endogeneity is difficult; if instrumental variable techniques are to be used, it would require variables correlated with changes in cohort size, but not independently correlated with schooling achievement.

4. Identification and Empirical Methodology

The prospects for identifying the effect of multi-grade teaching are better, even in regressions which include cohort-by-school fixed effects. This is because the decision to place a cohort in a multi-grade classroom represents a combination of a specific set of school-level attributes and a grade level decision, which compares the grade in question to that of other grades in the school along dimensions specific to the multi-grade decision. Because this relative measure changes as older cohorts exit the school and newer ones enter, the determinants of whether a cohort should be placed in a multi-grade classroom varies over time, even for the same cohort. We use the decision rules which guide the decision to place a cohort in a multi-grade classroom as instruments, using fixed effects to control for time-invariant attributes of the cohort and the school.

Our empirical analysis commences with results from two base models. The first utilizes fixed effects for different grades, and combines this with a set of school-level dummy variables. The second replaces school fixed effects with cohort-by-school fixed effects. We start our discussion of our approach to identification by describing these base models, before moving to a discussion of our instrumental variable strategy.

4.1 School fixed effects

Our base estimating equation is as follows:

$$(1) \quad Y_{icst} = \alpha_o + \alpha_1 M_{cst} + \alpha_2 X_{it} + \alpha_3 S_{st} + \mu_g + \mu_s + T_2 + \varepsilon_{icst}$$

For student i in cohort c of school s , Y_{icst} records his or her test score in year t . M_{cst} is an indicator variable for whether the child is in a multi-grade classroom, and is the variable of interest in this analysis. X_{it} is a vector of child characteristics – age and age squared, gender and caste. Time varying characteristics of the school such as total enrollment are included in S_{st} . The regression also includes a set of dummy variables for the grades included in the survey (μ_g), a set of school fixed effects (μ_s) and a dummy variable for year 2 of the survey (T_2). The residual error term (ϵ_{igst}) includes unobserved determinants of student achievement at the individual, grade and school level.

The inclusion of school fixed effects in (1) removes the influence of all time-invariant attributes of the school and community from the regression equation. The effect of multi-grade teaching is thus identified by variation in the types of classrooms within a school, as well as variation over time in the classroom environment. However, school choices regarding the placement of students in different classrooms remain a concern. It is likely that the decision to join students of one grade with those of another in a multi-grade environment represents specific attributes of the cohort in question, such as its size. If so, school fixed effects eliminate the bias caused by endogenous selection into classrooms and schools only if the variation in the size of cohorts across a school is random. This is unlikely to be the case when there is grade- and cohort-specific exit from schools, due either to drop out rates or transfers to private schools.⁴

⁴ Conversely, in economies where drop out rates are low and school choice is limited, one can make an argument that school fixed effects remove much of the bias in regressions of achievement on the attributes of a particular grade Lavy and Schlosser (2007), for example, use this strategy to evaluate gender peer effects using data for Israeli school.

4.2 *Cohort-by-school fixed effects*

A second set of estimates controls for time-invariant attributes of a given cohort in a given school by including a set of cohort- and school-specific fixed effects. We choose to incorporate cohort rather than grade fixed effects because, for reasons described below, the schooling environment in rural India generates less time-invariant variation in school quality within grades of a school than it does within cohorts.

One would observe time-invariant variation in the quality of schooling across grades within a school either because grade-specific curricula vary across schools or because key inputs such as teachers are assigned to specific grades and there is little re-assignment of teachers across grades once the initial assignment is made. Because of the relatively centralized nature of government schools in India, neither of these conditions applies. Curricula are dictated by authorities at the state, and sometimes national, level, so that, while there is grade-specific variation, it is uniform across schools. Consequently, variation in grade-level quality due to curricula or other aspects of school policy can be handled through a set of dummy variables for each grade, as included in equation (1).

Time invariant grade-specific variation in teacher quality is also of minimal concern. Decisions regarding teacher appointments in Indian government schools are made at the district level, with teachers appointed to a district cadre. Transfers of teachers across the schools in a district, as a matter of policy, are common and frequent. In the state of Andhra Pradesh, the site of this study, the official transfer policy makes a teacher eligible for transfer to a different school after having served only two years in any given school. Additionally, because schools are required to maintain a certain teacher-student ratio,

changes in enrollment also generate considerable movement in teachers across schools. This high mobility rate also suggests a limited role for time-invariant differences in quality across grades.

A more likely source of bias, then, is cohort differences in unobserved ability or other characteristics which directly influence learning and are correlated with the decision to place a cohort in a multi-grade environment. To control for this, we initially expand (1) to include cohort-by-school fixed effects, exploiting the availability of a panel of data spanning two years. Because the grades combined in multi-grade classrooms vary each year, we observe changes in this aspect of the classroom environment, for a given cohort in a school, over the two years of our study. This provides the basis for estimating the effects of multi-grade teaching, even in regressions which remove the effect of time-invariant characteristics of the cohort through cohort-by-school fixed effects.

However, mean ability and behavioral traits can change over time, as students exit the school or new students enter. And, even when these traits are invariant, the frequent change in teachers assigned to a school implies that the teacher quality experienced by a particular cohort can vary over time. It remains true that attributes of the cohort, such as cohort size, which determine multi-grade teaching but may also influence teacher allocations, can bias estimates of the effects of multi-grade teaching.⁵ Unless the endogeneity of the multi-grade decision is taken into account, estimates of the effect of a

⁵ The frequent changes in the teacher population in a school also implies that it is not possible to control for teacher quality by including teacher fixed effects; the frequent turnover of teachers in India's government schools generates a small set of teachers present in both survey years, who also switch from multi-grade classrooms to regular classrooms within that period.

multi-grade classroom on individual learning may be biased, even when we control for time-invariant characteristics of the cohort.

4.3 Instrumental variables

We therefore also report a set of results which use instrumental variables in conjunction with cohort-by-school fixed effects to eliminate any bias caused by the possible endogeneity of the allocation of cohorts to multi-grade classrooms. Our identification strategy builds on the recognition that the probability of observing a multi-grade classroom in a school is the outcome of a *joint* probability function: the probability at the school level of the decision to combine two or more grades in one classroom, or the probability that a school offers at least one multi-grade classroom, and the probability at the grade level of which grades to combine. Our approach is similar to the literature which exploits the discontinuity in the function relating class-size to enrollment to form instruments for class size. As in this literature, we build on the fact that the probability of observing a multi-grade classroom in a school is a specific non-linear function of school size. The strength of this strategy draws heavily on the fact that the specific non-linear relationship we exploit is unique to the multi-grade decision. The credibility of our estimates is enhanced because we combine the determinants of multi-grade teaching at the school level with grade-level features which determine which grades within a school should be combined in one classroom.

In the sections below, we separately discuss the determinants of multi-grade teaching at the school and cohort level.

A. *The probability of a school providing a multi-grade classroom*

The first building block in our instrumental variable strategy is the determination of the probability that a school has at least one multi-grade classroom. As stated above, our approach recognizes that the probability of observing a multi-grade classroom in a school is a specific non-linear step function of the number of teachers in the school, itself reflective of school size. If school size is sufficient to ensure 5 teachers, there will be no multi-grade teaching. With 4 teachers, at most 1 classroom will be a multi-grade one, so that the probability of a multi-grade classroom is 0.4. With 3 teachers, the probability of a multi-grade classroom is 0.8, and with two teachers, there will be no single-grade classrooms. This suggests the following predicting equation for the probability that a school has at least one multi-grade classroom:

$$(2) \quad \Pr(M_{gst} = 1) = d_{1st} + 0.8d_{2st} + 0.4d_{3st}$$

Where $d_{1st}=1$ if school s in year t has two or fewer teachers; 0 otherwise

$d_{2st}=1$ if school s in year t has 3 teachers; 0 otherwise

$d_{3st}=1$ if school s in year t has 4 teachers; 0 otherwise

Equation (2) suggests a regression discontinuity design; schools with more than 4 teachers will not have any multi-grade teaching so that a comparison of outcomes across “large” and “small” schools, as defined by this particular cut-off level, will identify the

effects of schools with multi-grade teaching to those without. We do better than this by exploiting the fact that even for schools with fewer than 5 teachers, the probability of observing a multi-grade classroom varies, in a known way, at specific size cut-off levels. It is unlikely that this unique non-linear function affects any other determinants of schooling, other than the multi-grade decision. It is certainly possible that the *total* number of teachers may affect learning⁶ or that certain other classifications of total teacher numbers may matter. For example, many believe that single teacher schools are particularly detrimental to learning, since teacher absence on any given day results in a complete absence of teachers in the school (Dreze and Sen 1995). Our function implies that it is a particular cut-off of 4 or less teachers (and the associated enrollment level) that matters and that, below this cut-off level, the implications of 1, 2, 3, or 4 teachers varies.

Rather than use the number of teachers in the school to calculate the d_i 's ($i=1-3$), we do so exploiting the rules which allocate teachers to schools on the basis of school enrollments. Teacher allocations in India are a combination of a “minimum rule,” which proscribes the minimum number of teachers in any given school, and a target student-teacher ratio of 40:1. Concerned about schools with less than two teachers, the Central Government provides funds to state governments to ensure that all schools have at least two teachers, and that schools with enrollments in excess of 100 have 3 teachers.⁷ At enrollment levels above 100, state governments are instructed to ensure a student teacher

⁶ Perhaps through its effect on a school-wide student-teacher ratio, though most would argue that it is the class student-teacher ratio that is relevant.

⁷ These rules were implemented as part of the Central Government's Operation Blackboard, introduced in 1986, and the expanded Operation Blackboard of 1992. Currently, funding for the provision of teachers at these levels of school enrollments is provided to state governments on a sharing basis with the Central Government, through its program *Sarva Shiksha Abhiyan*.

ratio of 40:1. The target is monitored at the school level, through the annual collection of school-level data on students and teachers.

This implies that the number of teachers is given by the following rule:

$$(3) \quad T_{st} = 2 \text{ if total enrollment } \leq 100 \\ = x \text{ if total enrollment } > 100$$

where x is the unique integer such that $x - 1 < \frac{\text{total enrollment}}{40} \leq x$.

This rule is therefore a modified version of Maimonides' rule, which proscribes that a new teacher be added once enrollments exceed 40, with subsequent additions as enrollments cross multiples of 40.⁸

Applying this rule to determine the d_i 's in equation (2) implies that the probability of observing a multi-grade classroom in a school is a particular step function of total enrollments. As previously argued with respect to the number of teachers in the school, it is unlikely that this unique non-linear function affects any other determinants of schooling, other than the multi-grade decision. *Total* enrollments, of course, influence the financial resources of the school. But, because we exploit the specific non-linearity embodied in the function relating the probability of observing a multi-grade classroom to the number of teachers (and hence total enrollments), it is possible to allow enrollments to directly influence learning, through the inclusion of a quadratic or higher-level polynomial of

⁸ This rule, implemented in several countries, was first used by Angrist and Lavy (1999), and then by others (Urquiola 2006) to identify the effects of classroom size on achievement.

enrollments in regression equations (we report results which incorporate a cubic function of enrollments).

To assess the validity of this rule in the allocation of teachers, figure 2 plots the actual and predicted number of teachers by total school enrollments, using data from a census of all rural schools in the 14 blocks of West Godavri districts where our survey was conducted. The graph suggests that equation (3) predicts well the number of teachers allocated to schools. Figure 3 plots the actual and predicted probability of observing a multi-grade classroom in a school as a function of total enrollment, with the latter determined by combining equation (2) and (3). Again, the predicted probability that a school has a multi-grade classroom closely tracks the observed probability.

The determination of which cohorts within a school should be in multi-grade rooms

Once it is known how many grades or cohorts within the school need to be taught in a multi-grade environment, the school authorities are responsible for determining which cohorts should be combined with others. While unobservable attributes of the cohort may play a role, headmasters inform us that a primary determinant of this decision is the need to ensure equality in class size. Thus, the cohort's *relative* size matters: smaller cohorts are much more likely to be taught in combination with another class, suggesting in turn the importance of a cohort's size rank in this decision.

Further, for any given cohort, the relative size of the cohort immediately younger to it is also important. From table 3, it is clear that there is a preference to combine younger cohorts in multi-grade classrooms, and that adjacent cohorts are much more likely to be

grouped together. This suggests that it is not just the size ranking of the cohort in question which is important, but that the rank of the cohort immediately younger may also be significant. The preference to place the youngest grades in multi-grade classrooms also suggest that a specific grade effect: students in grade 1 or 2 are more likely to be observed in multi-grade classrooms.

B. Instruments for multi-grade classrooms:

Since the probability of a cohort being taught in a multi-grade environment reflects the joint probability of observing a multi-grade classroom in a school as well as the decision of which cohorts to combine, we instrument M_{gst} with interactions of variables which reflect these two distinct probabilities. Let $Pr(mgrade)$ be the non-linear function of total enrollment determined by a combination of (2) and (3), which determines whether a school offers a multi-grade classroom. Our instrument set includes $Pr(mgrade)$ and the interaction of this variable with the size rank of the cohort relative to other cohorts in the school ($rank$), the relative rank of the younger cohort ($rank_yng$), the interaction of $rank$ and $rank_yng$, and the interaction of $rank$ with a dummy indicator for whether the cohort is in grade 2.

Of course, the relative size of a cohort within a school may also influence the allocation of teachers to classrooms. Our identification strategy exploits the co-existence of schools with multi-grade teaching and those without. We can therefore allow for any independent effect of the size rank of a cohort, and the one immediately below it, basing identification only on the interaction of rank with $Pr(mgrade)$. That is, similar to a

difference-in-difference strategy, we exploit the fact that size rankings have an additional effect on schooling attainment in multi-grade schools, because they determine which cohorts will be grouped together in the multi-grade classrooms present in the school. The regression framework, however, allows us to control for the endogeneity of the probability that a school has a multi-grade classroom, while also allowing controls for the direct (non-linear) effects of total enrollment and size rank on schooling attainment.

4.4 *Controlling for endogenous variation in cohort size over time*

In regressions with school fixed effects, as well as those with cohort-by-school fixed effects, the observed variation in $Pr(mgrade)$, the probability of observing a multi-grade classroom within a school, reflects variation in school size over our survey years. Some element of this variation is random, caused by random demographic differences in the size of entering and exiting cohorts; if there was no change in net enrollments for a given cohort over its school years, variation in school size would still occur because of differences in the size of the cohort exiting the school (upon completing 5th grade), and that newly entering the school in grade 1.

However, school size also varies because of changes in net enrollment rates over time for any given cohort, changes which could well be caused by time-varying cohort specific changes in school quality. That is, variation in school and cohort size may also include an endogenous component. Though our instruments exploit a specific non-linear relationship between multi-grade schools and school enrollments, it may still be the case

that changes in the number of multi-grade classrooms for a given school reflect endogenously induced changes in school enrollment.

We therefore fix the enrollment level of each cohort at an initial level, thereby eliminating any variation in cohort size over time. Variation in school size then comes completely because of differences in the fixed size of the cohort exiting in the school and that of the entering first grade cohort. Using lagged data on grade-level enrollments for the 2003-04 academic year, a cohort's size is set equal to its entry level size for the 3 cohorts who were in grade 1 in either 2003-04, 2004-05 or 2005-06. For the remaining cohorts (those in grades 4 and 5 in 2005-06), we set enrollments at their 2003-04 levels (when they were in grades 2 and 3 respectively).

Ideally, we would have liked data on first grade enrollments for all cohorts. Without this, the comparison in the size of the cohort currently in first grade and that in fifth grade is based on data for the fifth grade class when it was in 3rd grade, and may include some endogenously induced variation in size due to the quality of schooling experienced by this cohort in first and second grade. We feel that such bias may be minimal, primarily because drop out rates, as well as transfers across schools, occur primarily at higher grades, notably in the fifth grade, before enrollment in middle school, but also in fourth grade.⁹

⁹ Systematic grade-level variation in drop out rates as a student ages are captured by the dummy variables for different grades included in our regression. Drop out is highest in the transition from primary to middle school. An additional concern is transfers from one school to another. While there are few instances of transfers from one government school to another, there is significant movement, from private schools to government schools. This occurs primarily in the 5th grade, in a phenomenon that is commonly referred to as 5th grade transfer. Private schools predominate at the primary level, and a majority of them are "unrecognized" schools which operate without completing the required government formalities. The higher infrastructural costs associated with operating middle and high schools result in very few unrecognized schools at this level. Movement into a recognized school requires a transfer certificate, which is not available

To further reduce the possibility for bias from this source, we estimate a “value added” specification, which conditions on the student’s initial test score. For this purpose, we exploit the fact that we have data on multiple tests for a student in any given grade, including an initial test, conducted at the start of the year, and those from half-yearly examinations. Our regressions are based on the results reported in half-yearly examinations, and we use the test scores from initial tests conducted at the start of the school year in the value added regressions. Conditioning on initial achievement levels at the start of the school year controls for the effects of historical determinants of school quality, including cohort-specific effects experienced in early schooling years.

We use the same fixed enrollment levels for each cohort to calculate its size rank amongst existing cohorts in the school. This implies that the variation in a particular cohort’s size rank similarly reflects variation only in the size of the outgoing and incoming class, and does not vary endogenously.

4.5 *Interpreting the Effect of Multi-grade classrooms*

The primary reason one would expect students placed in a multi-grade classroom to do worse than those in single-grade classrooms is that the teacher has to divide his or her time between the teaching of two or more distinct syllabi. As a consequence, instructional time will be reduced. The extent of the reduction will depend on how many grades are combined in one classroom; the negative effect of combining three grades will exceed that of combining two. As a consequence, the coefficient on multi-grade teaching will likely

from an unrecognized school. Therefore, many students enrolled in private unrecognized schools transfer to government schools in the fifth grade, to ensure their eligibility for formal middle schools.

vary depending on school and cohort size, the factors which determine how many grades are to be combined in one classroom.

Additionally, if schooling attainment is affected by attributes such as classroom size and the caste composition of students, then the change in these attributes caused by multi-grade teaching will also be reflected in the coefficient on a multi-grade classroom. To see this, consider the effect of class size. Let S_c represent the size of cohort c . Let CL_c denote the size of the class, if the school were required to place cohort c in a multi-grade classroom. If M_c is, as before, the indicator for a multi-grade classroom, then the observed class size for cohort c , CL_c^* , is then determined by the following equation:

$$(4) \quad CL_c^* = CL_c * M_c + S_c * (1 - M_c) = (CL_c - S_c) * M_c + S_c$$

If classroom size, or other attributes of the classroom such as student composition, affects learning, then the coefficient on multi-grade classrooms from regressions which omit these additional attributes will combine the “pure” learning disadvantage of multi-grade teaching with an additional component reflecting the change in classroom attributes caused by combining different grades in one classroom. This, again, suggests a random coefficient specification, with the effect of multi-grade classrooms varying across cohorts in a school.

As noted by Heckman and Robb (1985), IV estimation generates biased coefficients in a random coefficient model, since the regression equation identifies the mean value of the coefficient, $(\bar{\alpha})$, while the error term includes the interaction of the deviation from the mean, $\varepsilon = \alpha - \bar{\alpha}$, with the indicator of multi-grade teaching, M_c .

Instruments for M_c will therefore be correlated with the regression error. However, control function estimators avoid this problem and produce consistent estimates, even in random coefficient models. Therefore, while we provide a set of results using instrumental variables, we also present results which utilize control functions to identify the effects of multi-grade teaching. The next section briefly describes our control function estimators.

4.6 Control Function Estimators

Adopting an index function framework, let IN_c be a function which defines the decision to place a cohort in a multi-grade classroom (for notational simplicity, we suppress the dependence on school, s , and time, t):

$$(5) \quad IN_c = Z_c \gamma + V_c$$

Assuming a random coefficient model, let the coefficient on multi-grade teaching be $\alpha_c = \bar{\alpha}_c + \varepsilon_c$. The regression equation (ignoring other conditioning variables) is:

$$(6) \quad E(Y_{ic} | M_c, Z_c) = \alpha_o + M_c \bar{\alpha}_c + E(U_{ic} + M_c \varepsilon_c | M_c, Z_c)$$

If the functional form of $E(U_{ic} + M_c \varepsilon_c | M_c, Z_c)$ is known up to a finite vector, then $\bar{\alpha}_c$ can be consistently estimated under a standard set of assumptions (outlined in Heckman and Robb 1985).

We adopt the standard assumption that (U_{ic}, V_c) are joint normal random variables with zero mean and a finite nondegenerate covariance matrix. Writing (6) as:

$$(7) \quad E(Y_{ic} | M_c, Z) = \alpha_o + M_c \bar{\alpha}_c + M_c E(U_{ic} + M_c \varepsilon_c | M_c = 1, Z_c) \\ + (1 - M_c) E(U_{ic} + M_c \varepsilon_c | M_c = 0, Z_c)$$

and defining:

$$\lambda_{1c} = \frac{\varphi(Z_c \gamma)}{\Phi(Z_c \gamma)}; \quad \lambda_{2c} = \frac{\varphi(Z_c \gamma)}{\Phi(-Z_c \gamma)}$$

We estimate (7), replacing the two conditional expectations on the right hand side of (7) with λ_{1c} and λ_{2c} , respectively. The inclusion of previously estimated variables in the regression equation requires an adjustment of standard errors to allow for estimation error. This is done through a standard δ -method correction (Murphy and Topel 1985).

As a robustness check, we also report results from a weaker functional form assumption, allowing $E(U_{ic} + M_c \varepsilon_c | M_c, Z_c)$ to be a function of $\Pr(M_c | Z_c)$. Assuming that this latter probability is normal, we define the following control functions:

$$K_{1c} = \Phi(-Z_c \gamma) + (\Phi(-Z_c \gamma))^2 \quad \text{and} \quad K_{2c} = \Phi(Z_c \gamma) + (\Phi(Z_c \gamma))^2$$

and report an alternative set of results, replacing λ_{1c} , λ_{2c} with K_{1c} , K_{2c} . Either methodology is preferable to the use of the propensity function as a control function, since the latter is

applicable only when the selection into multi-grade classrooms is based exclusively on observables.

4.7 Comparing the effects of multi-grade teaching, class size and caste composition

The recognition that the variation in classroom size and other classroom attributes partly emanates from the institution of multi-grade teaching suggests a new methodology for identifying their effects on schooling attainment, and comparing them to those of multi-grade teaching. As previous noted, while non-experimental studies of class size effects generally use the rules which determine the division of grades into classrooms as a source of identification, this does not get around the problem of cohort-specific factors which may bias the effects of class size. This is also true of research on the effects of the racial and ethnic composition of the student body on learning. The most credible estimates exploit variation across adjoining cohorts within a school, through the inclusion of grade-by-school fixed effects in regressions. Even here, the potential for bias exists because variation in the size of adjacent cohorts enrolled in any given school reflects both endogenous and exogenous sources of variation.

The existence of multi-grade classrooms provides a unique opportunity to identify the effects of classroom size and the caste composition of students in a classroom on individual learning. Because of multi-grade teaching, the size and composition of a classroom changes, even for a cohort which experiences no change in its own net enrollment rate; the variation in the classroom environment is induced purely by the decision to combine the cohort in question with a different cohort in one classroom. That

is, the change in the classroom environment is purely a consequence of multi-grade schooling.

To ensure that this is the case, we continue to fix the size of each cohort at its initial level, so that it does not vary over our survey years. Once the multi-grade decision is made, the variation in classroom size and the caste composition of students is now exogenous. Because of this, the control function estimator yields consistent estimates of the effect of classroom attributes, as long as the conditioning removes any correlation between classroom attributes and the error term in the regression function. This condition is trivially met in this case, since the variation in classroom attributes such as size and caste composition experienced by any given cohort occurs only as a consequence of the decision to combine it with another cohort in one classroom.

5. Results

5.1 OLS Regressions

We start the empirical analysis of this paper with OLS regressions of language and math scores on an indicator of whether the child was in a multi-grade classroom. All regressions are of the “value added” form, including test scores from base tests conducted at the start of the academic year. The regressions also include a set of child-specific covariates (caste, age, squared age and gender), a cubic in initial school enrollment,¹⁰ a dummy variable for the second survey year, and grade level fixed effects. The two regressions reported in table 4 (for each subject) differ in that the first includes only a set of school fixed effects, while

¹⁰ Calculated by fixing the size of each of the 5 cohorts in the school at their entry level, as previously described.

the second includes cohort-by-school fixed effects. As noted, the current literature on the effect of classroom attributes on learning draws heavily on a similar methodological approach, exploiting variation across cohorts within a school in the variable of interest.

The regressions reveal a small but significant negative effect from multi-grade teaching for language scores. For mathematics, though the effect is negative, it is small and statistically insignificant in regressions which include cohort-by-school fixed effects. Using the results from regressions which include cohort-by-school fixed effects, the implied elasticity of scores with respect to multi-grade teaching is 0.006 for language and 0.001 for mathematics.

These regressions also reveal a negative coefficient on the indicator variable for scheduled caste; for any given cohort in any given school, children from scheduled castes and tribes do worse than children from upper castes. Because these regressions control for all time-invariant attributes of the school (and hence the community), as well as that of the cohort, the negative coefficient on the SC/ST indicator does not reflect that component of schooling inequalities which results from (time-invariant) differences in the quality of schools attended by SC/ST students relative to others. That is, it does not reflect the “between” school component of schooling inequality which, given the significant differences between hamlet and main village schools (table 2), is likely to be large.

5.2 *First Stage and Reduced Form Regressions*

Before discussing results from IV and control function estimation, table 5 presents results from first stage regressions of multi-grade teaching on the set of instruments. Our

instrument set comprises the interaction of $Pr(mgrade)$ with the cohort-specific variables which determine whether the cohort in question, within those in a school, will be placed in a multi-grade classroom. As previously argued, these cohort-level variables may independently affect learning through, perhaps, their influence on teacher assignments and the distribution of school resources across grades. Thus, they are included as regressors in all regressions, and excluded from the instrument set.

Our instruments affect the probability of a multi-grade classroom in the expected direction, providing support for their validity. The non-linear function of (initial) school enrollment used to predict multi-grade teaching ($Pr(mgrade)$) has a statistically significant positive effect on the probability that a child is in a multi-grade classroom. Moreover, as expected, interactions with the size rank of the cohort (with low rank implying small size) generates a negative effect. Within schools which are likely to have multi-grade classrooms, larger cohorts are less likely to be placed in a multi-grade classroom. Because there is a preference for grouping adjoining cohorts together, this is true also if the cohort younger to the one in question is relatively large. The regression estimates also suggest that the negative effect of own size is tempered if the younger cohort is also large. Finally, amongst schools likely to have multi-grade teaching, lower grades are more likely to be placed in multi-grade classrooms.

Because of the non-linear instrumental variable strategy of this paper, it is not possible to test the validity of our exclusion restrictions through standard over-identification tests. However, reduced form regressions of language and math scores on the full set of instruments provides some support for our identifying assumptions. If our

instrument set affects test scores only through their effect on multi-grade teaching, their coefficients in reduced form regressions should be opposite in sign to those in the first stage regression of instruments on multi-grade teaching. The second and third regressions reported in table 5 confirm that this is the case, lending support to our identifying strategy.

5.3 *Instrumental Variables and Control Function Estimates*

Results from IV and control function regressions are reported in tables 6 and 7 respectively. In both cases, we report results which use school fixed effects as well as our preferred estimates which incorporate cohort-by-school fixed effects.

The IV estimates are generally larger than the control function estimates, and both are significantly larger than OLS estimates (a discussion of the economic magnitude of the effect of multi-grade teaching is deferred to the next section). This suggests that attributes of the cohort which affect placement in multi-grade classrooms, but also directly affect learning, negatively bias OLS estimates of the consequences of multi-grade teaching. For example, it may be that better behaved or smarter cohorts are placed into multi-grade classrooms, or that more capable teachers are assigned to these classrooms. Failure to deal with the endogenous placement of cohorts and teachers into multi-grade classrooms generates lower estimates of their negative effect on learning.

Statistical confirmation that the IV estimates differ from those obtained from OLS regressions comes from the significant coefficient on the control functions ($\lambda_1 * M, \lambda_2 * (1 - M)$) in the results reported in table 7. For the preferred set of regressions

which utilize cohort-by-school fixed effects (regressions 2 and 5), the control functions are statistically significant at conventional levels.

Regressions (3) and (6) in table 7 explore the sensitivity of results to the functional form assumptions which underlie the construction of our initial choice of control functions. Replacing λ_1 and λ_2 with a quadratic in the estimated probability of the cohort being in a multi-grade classroom generates generally similar regression results.

The estimates reported in regressions (2) and (5) of table 7 imply an elasticity for both language and math tests scores with respect to multi-grade teaching of 0.05. In the last section of this paper, we explore what this implies for caste-based schooling inequalities. Prior to that, we compare the effect of multi-grade teaching to that of the caste composition of the classroom, and of classroom size.

5.4 Comparing the effect of multi-grade teaching, caste composition and classroom size

As discussed in the previous section, we are able to identify the effect of the caste composition of the classroom and of classroom size, controlling for the endogeneity normally associated with these attributes. We do this by fixing the size and caste composition of each cohort at its initial level, allowing the variation in these attributes to come only because of changes in the multi-grade environment for a given cohort in the two years of our survey. In this context, the use of control functions to eliminate the bias introduced by the endogenous placement of cohorts in multi-grade classrooms simultaneously produces consistent estimates of the effect of classroom composition and size.

Table 8 reports results from regressions which include the proportion of SC/ST students in the classroom, as well as classroom sizes. A second regression (for each subject) includes the square of classroom size. This term was found to be statistically insignificant. Consequently, our discussion focuses on the estimates generated by the first regression for each subject.

Including additional classroom attributes reduces the coefficient on multi-grade teaching, confirming our earlier conjecture that the “overall” coefficient on multi-grade teaching revealed in previous regressions is partly a consequence of differences in multi-grade versus single-grade classrooms in terms of caste composition and size. While the reduction is small for language (so that the implied elasticity of -0.05 is unaffected), the results of this table suggest that the elasticity of math scores with respect to multi-grade teaching is lower, at -0.03.

The effect of the caste composition of the classroom differs across the teaching of language and mathematics. It does not significantly affect language scores. However, students placed in classrooms with a greater percentage of SC/ST students perform significantly poorly in mathematics; the elasticity of mathematics scores with respect to caste composition, at -0.05, exceeds that associated with multi-grade teaching (-0.03).

Finally, classroom sizes appear to be a more important determinant of learning than both multi-grade teaching and the caste composition of the classroom, particularly for the teaching of mathematics. For language, the implied elasticity of classroom size is comparable to that of multi-grade teaching (-0.06 compared to -0.05). For mathematics, however, a 1% increase in classroom size reduces mathematics scores by 0.12%.

6. Explaining Caste-based Schooling Differentials

The significant effect of classroom attributes on test scores suggests that differences in the quality of schools attended by scheduled caste and upper caste children may be an important factor explaining caste-based schooling inequalities. Data which reveals significant differences, along these dimensions, across the schools attended by children of different castes, would confirm this hypothesis.

For this purpose, we restrict our attention to the villages in which we surveyed both the main village and the hamlet schools, dropping villages in which we only surveyed the main village school. For this smaller sample, we compare classroom attributes for SC/ST students in the sample to those of upper caste students, regardless of the school they attended. This comparison thus provides estimates of the differences in the quality of classrooms attended by scheduled caste and other caste students.

For this sample of 15 villages, the data reveal that 47% of SC/ST students, but only 18% of upper caste students are enrolled in a multi-grade classroom. Confirming extensive segregation by castes in rural schools, SC/ST children attend schools where the proportion of SC/ST children in the classroom is 0.66. In contrast, the proportion of SC/ST students in the classroom is only 0.15 for upper caste children. However, there is little significant difference in the size of classrooms attended by children from scheduled castes and tribes, relative to their more privileged counterparts. Average classroom size for SC/ST students is 38.4, while it is slightly larger, at 39.7, for upper caste students.

The data therefore suggest that though classroom size is a more important determinant of learning than multi-grade teaching or the caste composition of the

classroom, caste based schooling differentials can only be explained by the latter two variables. These two factors play an approximately equal role in explaining caste-based differences in mathematics scores. For language, however, the primary school-level factor explaining caste differences in achievement is multi-grade teaching.

It is difficult to gauge how much of the overall difference in achievement (as measured by test scores) across castes is explained by multi-grade teaching. This is primarily because of the lack of estimates of the caste gap in achievement. The data used in this study also provide no information on the overall gap, because the test results reported do not come from standardized tests. While the centralization of testing at the level of the block and the district ensure that the tests administered to different schools have a similar level of difficulty, it is possible that teachers inflate test scores from poorly performing schools. Indeed, the relatively high average tests scores in hamlet schools (table 2) support this supposition. As a consequence, we are not able to ascertain the relative importance of school quality in explaining the overall difference in schooling attainment by caste.

We can, however, use our estimates to examine the magnitude of the difference in test scores which can be attributed to multi-grade teaching alone. For this purpose, we use the regression results from table 8 (the first regressions reported for each subject). We assume that scheduled caste students will, on average, face one year of multi-grade teaching, while the average upper caste student will only experience single-grade teaching. This assumption comes from the data on 15 villages which reveal that the total number of teachers in schools attended by scheduled caste students is, on average, 3.89, while it is 5.32 for upper caste students. To isolate the effects of multi-grade teaching, we assume that

all other determinants of schooling are equal, so that the test scores of SC/ST students are identical to those of upper caste students at the start of second grade. Starting at the same initial test score (60.4 for Language and 61.6 for mathematics), we assume that scheduled caste students are placed in a multi-grade classroom in grade 2, but in single-grade classrooms in the remaining grades. Upper caste students are assumed to be in single-grade classrooms throughout primary school.

We trace what this means for test scores over the course of the student's time in primary school, allowing test scores at the end of second grade to fall as a consequence of multi-grade teaching (using the regression estimates in table 8 to calculate the extent of this fall). For subsequent years, when both SC/ST and upper caste students are in single-grade classrooms, differences in learning arise only through the dynamic effect of initial test scores. That is, the lower learning in second grade implies a lower base at the start of third grade, and so on. The regression estimates of table 8 suggest regression to the mean, with the effect of (any) initial differences in test scores being reduced over time. We examine the extent of the gap in learning at the end of 5th grade caused by the placement of otherwise identical students in a multi-grade classroom in grade 2.

This calculation suggests a large gap in test scores, for both language and mathematics, at the end of the second grade. For mathematics, test scores for SC/ST students fall to 55.4, while those of upper caste students increase to 64.9. Regression to the mean implies that this gap is reduced over the years, but it remains sizeable at the end of fifth grade, when average test scores for SC/ST students is 66.8, and that of upper caste students is 69.2. Similarly for language scores, the initial gap of 12.5 points (52.0 for

SC/ST students versus 64.5 for upper castes) at the end of second grade is large enough to imply a significant gap at the end of fifth grade (of 4.7 percentage points, 69.0 for SC/ST students and 73.7 for upper castes), despite regression to the mean in test scores.

7. Conclusion

Despite the very small school sizes observed in rural India, there has been almost no attention paid in the economics literature to its potential effects on learning or on schooling inequalities. Small school sizes reflect rural India's residential structure and the school location policies of the Government of India. Indian villages are sub-divided, generally by caste, into small, geographically dispersed habitations, with those habited by scheduled castes and tribes being significantly smaller than others. In order to ensure easy access to schools, the Government adopted a policy which recognized the habitation as the relevant unit for school mapping exercises, and required schools to be available in each habitation. The small size of habitations and variation in this size has resulted in a school system which features a large number of very small schools, with those located in scheduled caste and tribe habitations being notably smaller than those in other habitations.

While small school size can affect school quality in many ways, this paper focuses on its consequences for multi-grade teaching. When schools have fewer than 5 teachers, they are forced to combine several grades in one classroom. We show that this reduces learning. And, because children of scheduled castes and tribes, who attend smaller schools, are more likely to experience multi-grade teaching, we find that it is a significant cause of caste-based schooling inequalities.

Extending the current literature on the effects of classroom attributes on schooling, we not only exploit the availability of panel data to aid in the identification of classroom effects, but also further reduce the potential for bias caused by time-varying changes in cohort quality by using an instrumental variable approach. Our approach exploits the discontinuity in the relationship between school size and the probability of observing multi-grade teaching in a school, using this to develop a difference-type estimator. Identification of classroom effects is achieved by combining this school-level determinant of multi-grade teaching with grade-level determinants, which recognize that the choice of which grades to combine in a single classroom reflects their size rank.

Recognizing that the existence of multi-grade classrooms introduces a new source of variation in class size and the caste composition of the student body, not generally present in countries with larger school sizes where multi-grade teaching is rare, we exploit this variation to generate new estimates of the effect of classroom size and student composition on student achievement. The estimates we derive are free of bias as a consequence of variation in unobserved attributes of different cohorts, or even of the same cohort (in any given school) over time. They therefore improve on the estimates currently available in the literature. We find that both classroom size and the caste composition of student matters for learning (though the latter matters only for mathematics scores, not for language), but that it is caste composition and multi-grade teaching which affect schooling inequality.

Table 1: Summary Statistics on Habitations in India, Survey Districts and Survey Villages

	India	Survey Districts	Survey Villages
Number of villages	586,986	2,153	60
Average number of habitations per village	2.1	2.0	4.0
Prop. SC habitations to total	0.38	0.38	0.25
Average population size, non-SC habitations	758	2,088	3,696
Average habitation size, SC habitations	425	607	1,120
<i>% of habitations with pop <300</i>			
All	44%	24%	--
SC/ST	55%	46%	--
<i>% of habitations with pop <500</i>			
All	62%	37%	--
SC/ST	74%	64%	--

Note: Data for India and Survey Districts are from the 7th All India Education Survey. Data for survey villages is survey data.

Table 2: Summary statistics on Main Village (non-SC) and Hamlet (SC/ST) schools in survey villages

	Main Village School (non SC)	Hamlet School (SC/ST)
Average Enrollment (2005-06)	153.93 (60.58)	82.94 (53.45)
Prop. SC/ST	0.17 (0.20)	0.68 (0.32)
Average Language score (August 2004)	62.60	67.19
Average Math score (August 2004)	63.83	66.28
Total number of teachers	5.09 (2.24)	2.63 (2.06)
Prop. of schools with at least one multi-grade class	0.39 (0.49)	0.82 (0.40)
Prop. of schools with 2 multi-grade classes	0.28 (0.46)	0.69 (0.48)
Prop. teachers SC/ST	0.21 (0.26)	0.31 (0.35)
Prop. teachers male	0.53 (0.27)	0.60 (0.38)
Teachers' mean age	37.24 (4.66)	37.28 (5.04)
Teachers' mean yrs teaching exp.	10.92 (4.81)	9.97 (5.32)
Prop. teachers with B.Ed or higher	0.74 (0.24)	0.70 (0.30)
Prop. teachers B.A. or higher	0.83 (0.22)	0.69 (0.31)

Note: statistics are calculated from survey data on 45 main village schools and 15 hamlet schools. Figures in brackets are standard errors. Average test scores are from data on test scores of children in grades 2-4 (sample size=8635 in main village schools and 1553 in hamlet schools).

Table 3: Probability of different grades being combined in multi-grade classrooms

Probability of being combined in multi-grade classroom with					
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Grade 1	1	0.36 (0.48)	0.10 (0.30)	0.08 (0.26)	0.08 (0.28)
Grade 2		1	0.13 (0.34)	0.11 (0.31)	0.03 (0.16)
Grade 3			1	0.13 (0.33)	0.07 (0.25)
Grade 4				1	0.13 (0.34)
Grade 5					1

Source: Survey data

Table 4: OLS Estimates of the Effect of Multi-grade teaching

	Language (Telugu)		Mathematics	
Multi-grade classroom	-1.97*	-1.52*	-0.94 ⁺	-0.36
	(0.54)	(0.77)	(0.57)	(0.74)
Initial test results	0.71*	0.72*	0.63*	0.63*
	(0.01)	(0.007)	(0.01)	(0.008)
SC/ST	-0.98*	-0.93*	-1.52*	-1.56*
	(0.45)	(0.48)	(0.45)	(0.46)
Age	-0.11 ⁺	-0.09	-0.19*	-0.65*
	(0.06)	(0.28)	(0.06)	(0.26)
Age square	0.0001 ⁺	0.0001	0.0001*	0.006*
	(0.00003)	(0.003)	(0.00003)	(0.003)
Male	-0.91*	-0.98*	-0.56*	-0.45 ⁺
	(0.27)	(0.28)	(0.27)	(0.28)
School enrollment	-0.83*	-0.77*	-1.08*	-1.17*
	(0.14)	(0.14)	(0.14)	(0.14)
(School enrollment) ²	0.003*	0.003*	0.004*	0.004*
	(0.001)	(0.001)	(0.001)	(0.001)
(School enrollment) ³	-3.1 e-6*	-3.18 e-6*	-4.5 e-6*	-5.47 e-6*
	(1.02 e-6)	(1.07 e-6)	(1.1 e-6)	(1.04 e-6)
Grade fixed effects	Yes	Yes	Yes	Yes
School fixed effects	Yes	No	Yes	No
Cohort-school fixed effects	No	Yes	No	Yes
Sample size	11,038	9,832	11,037	9,835
Regression F [@]	558.14	115.68	284.70	98.11
	(0.00)	(0.00)	(0.00)	(0.00)

Note: * Significant at 5% level

⁺ Significant at 10% level

[@] Figures in brackets are probability greater than F

Table 5: First Stage Regressions

	Multi-grade classroom	Language test scores	Math test scores
Pr (mgrade)	0.68* (0.06)	-16.95* (4.82)	-23.94* (4.59)
Pr (mgrade)	-0.22* (0.02)	4.91* (1.10)	4.87* (1.12)
* size rank			
* Rank younger classroom	-0.13* (0.02)	2.21+ (1.23)	2.47* (1.20)
* Size Rank * rank younger classroom	0.07* (0.005)	-0.84* (0.33)	-0.65* (0.34)
* Size rank * grade 2	0.03* (0.01)	-1.20* (0.41)	-3.02* (0.44)
Size rank	0.01* (0.007)	-2.93* (0.57)	-4.20* (0.57)
Rank younger classroom	0.02* (0.007)	0.51 (0.59)	-1.36* (0.57)
Rank * rank younger classroom	-0.002+ (0.002)	0.50* (0.16)	0.83* (0.16)
School enrollment	0.004+ (0.002)	-0.67* (0.17)	-1.07* (0.17)
(School enrollment) ²	-9.6 e-6 (0.00001)	0.003* (0.001)	0.004* (0.001)
(School enrollment) ³	-2.4 e-9 (1.5 e-8)	-2.85 e-6* (1.13 e-6)	-4.25 e-6* (1.11 e-6)
Initial test score	-0.0001 (0.0001)	0.72* (0.01)	0.63* (0.01)
Grade fixed effects	Yes	Yes	Yes
Cohort-school fixed effects	Yes	Yes	Yes
Sample size	9832	9,867	9,870
Regression F [@]	227.25 (0.00)	111.48 (0.00)	95.94 (0.00)
F statistic on instruments [@]		7.45 (0.00)	18.76 (0.00)

Note: * Significant at 5% level

+ Significant at 10% level

@ Probability greater than F

Table 6: IV regressions

	Language (Telugu)		Mathematics	
Multi-grade classroom	-5.23*	-17.09*	-12.65*	-19.27*
	(2.08)	(3.76)	(2.16)	(3.94)
Initial test score	0.70*	0.72*	0.63*	0.62*
	(0.007)	(0.01)	(0.01)	(0.01)
SC/ST	-1.00*	-0.89 ⁺	-1.55*	-1.67*
	(0.45)	(0.48)	(0.46)	(0.49)
School enrollment	-0.92*	-0.92*	-1.20*	-1.28*
	(0.14)	(0.15)	(0.14)	(0.15)
(School enrollment) ²	0.003*	0.004*	0.004*	0.005*
	(0.001)	(0.001)	(0.001)	(0.001)
(School enrollment) ³	-3.7 e-6*	-4.46 e-6*	-5.06 e-6*	-6.05 e-6*
	(1.04 e-6)	(1.10 e-6)	(1.08 e-6)	(1.08 e-6)
Size rank	0.57*	-2.49*	-0.69*	-3.96*
	(0.29)	(0.50)	(0.30)	(0.52)
Rank younger class	0.93*	0.82 ⁺	0.01	-1.33*
	(0.32)	(0.47)	(0.32)	(0.48)
Rank*rank younger class	-0.25*	0.55*	0.14	1.03*
	(0.09)	(0.15)	(0.19)	(0.15)
Grade fixed effect	Yes	Yes	Yes	Yes
School fixed effects	Yes	No	Yes	No
Cohort by school fixed effects	No	Yes	No	Yes
Sample size	11,034	9,832	11,033	9,836
Regression F [@]	533.33	111.14	268.84	74.03
	(0.00)	(0.00)	(0.00)	(0.00)

Note: * Significant at 5% level

+ Significant at 10% level

@ Figures in brackets are probability greater than F

Table 7: Control Function Estimates

	Language			Mathematics		
	(1)	(2)	(3) [@]	(4)	(5)	(6) [@]
Multi-grade classroom	-4.32* (1.46)	-13.98* (2.62)	-10.07* (2.34)	-7.06* (1.42)	-13.19* (2.60)	-9.95* (2.33)
Initial test score	0.70* (0.01)	0.72* (0.01)	0.72* (0.01)	0.63* (0.01)	0.63* (0.01)	0.63* (0.01)
SC/ST	-0.99* (0.45)	-0.91* (0.47)	-0.91* (0.47)	-1.50* (0.45)	-1.52* (0.46)	-1.54* (0.46)
School enrollment	-0.95* (0.14)	-1.08* (0.15)	-1.02* (0.15)	-1.24* (0.14)	-1.43* (0.15)	-1.38* (0.15)
(School enrollment) ²	0.004* (0.001)	0.004* (0.001)	0.004* (0.001)	0.005* (0.001)	0.005* (0.001)	0.005* (0.001)
(School enrollment) ³	-3.9 e-6* (1.04 e-6)	-5.12 e-6* (1.10 e-6)	-4.86 e-6* (1.1 e-6)	-5.3 e-6* (1.1 e-6)	-6.57 e-6* (1.1 e-6)	-6.4 e-6* (1.1 e-6)
Size rank	0.59* (0.28)	-2.49* (0.48)	-2.23* (0.47)	-0.43* (0.28)	-3.77* (0.49)	-3.57* (0.48)
Rank younger class	0.95* (0.31)	0.59 (0.47)	0.77 ⁺ (0.47)	0.18 (0.31)	-1.50* (0.48)	-1.35* (0.48)
Rank*rank younger class	-0.26* (0.09)	0.48* (0.14)	0.42* (0.13)	0.07 (0.09)	0.91* (0.14)	0.87* (0.14)
$\Lambda_{1c} * M_c$	1.72 ⁺ (1.02)	6.72* (1.66)	4.57* (1.58)	4.66* (0.98)	6.65* (1.60)	5.17* (1.54)
$\Lambda_{2c} * (1 - M_c)$	-1.45 (1.01)	-7.13* (1.77)	-5.16* (1.63)	-2.86* (0.93)	-7.90* (1.73)	-6.06* (1.59)
School fixed effects	Yes	No	No	Yes	No	No
Cohort by school fixed effects	No	Yes	Yes	No	Yes	Yes
Sample size	11,034	9,832	9,832	11,025	9,836	9,836
Regression F	522.00 (0.00)	111.14 (0.00)	112.55	268.39 (0.00)	96.70 (0.00)	96.71 (0.00)

Note: All regressions include grade fixed effects.

[@]Regression results in columns 3 and 6 use the alternative control functions, κ_1 and κ_2 , based on the probability of observing a multi-grade classroom, as explained in the text

Table 8: Comparing Effects of Multi-grade teaching to Class Size and Student Caste Composition

	Language (Telugu)		Mathematics	
Multi-grade classroom	-12.56 [*] (2.66)	-12.60 [*] (2.82)	-9.47 [*] (2.66)	-9.52 [*] (2.79)
Prop. SC/ST in classroom	-0.63 (5.58)	-0.61 (5.56)	-14.53 [*] (6.04)	-14.50 [*] (6.03)
Class size	-0.09 [*] (0.04)	-0.09 (0.11)	-0.19 [*] (0.04)	-0.19 ⁺ (0.11)
Class size square	--	-0.00004 (0.001)	--	-0.0001 (0.001)
Initial test score	0.72 [*] (0.01)	0.72 [*] (0.01)	0.63 [*] (0.01)	0.63 [*] (0.01)
SC/ST	-0.90 ⁺ (0.47)	-0.90 [*] (0.47)	-1.53 [*] (0.46)	-1.53 [*] (0.46)
School enrollment	-1.03 [*] (0.15)	-1.03 [*] (0.15)	-1.31 [*] (0.15)	-1.31 [*] (0.15)
(School enrollment) ²	0.004 [*] (0.001)	0.004 [*] (0.001)	0.005 [*] (0.001)	0.005 [*] (0.001)
(School enrollment) ³	-4.77 e-6 [*] (1.11 e-6)	-4.78 e-6 [*] (1.11 e-6)	-5.85 e-6 [*] (1.07 e-6)	-5.85 e-6 [*] (1.07 e-6)
Size rank	-2.50 [*] (0.48)	-2.50 [*] (0.48)	-3.78 [*] (0.49)	-3.78 [*] (0.49)
Rank younger class	0.62 (0.47)	0.62 (0.47)	-1.42 [*] (0.48)	-1.42 [*] (0.48)
Rank*rank younger class	0.47 [*] (0.14)	0.47 [*] (0.14)	0.91 [*] (0.14)	0.91 [*] (0.14)
$\Lambda_{1c} * M_c$	7.35 [*] (1.71)	7.36 [*] (1.74)	7.52 [*] (1.64)	7.53 [*] (1.66)
$\Lambda_{2c} * (1 - M_c)$	-7.78 [*] (1.79)	-7.77 [*] (1.79)	-8.95 [*] (1.75)	-8.94 [*] (1.76)
Grade fixed effects	Yes	Yes	Yes	Yes
Cohort by school fixed effects	Yes	Yes	Yes	Yes
Sample size	9,832	9,832	9,836	9,836
Regression F [@]	112.13 (0.00)	111.66 (0.00)	96.10 (0.00)	95.65 (0.00)

Note: ^{*} Significant at 5% level

⁺ Significant at 10% level

[@] Figures in brackets are probability greater than F

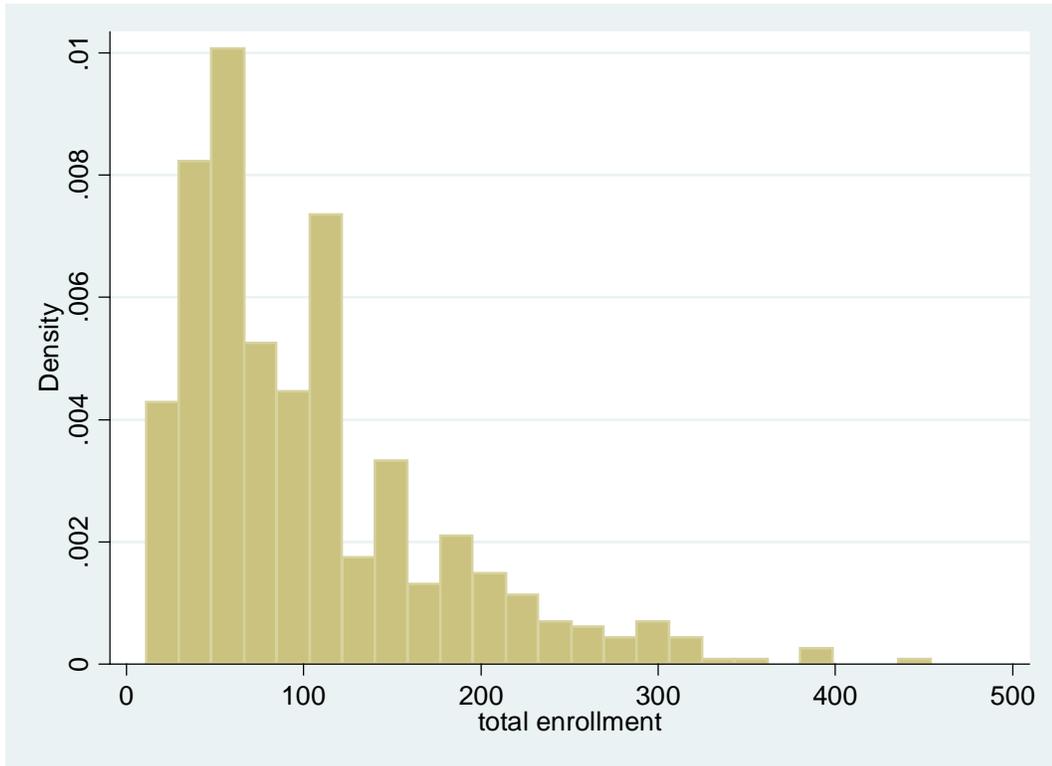


Figure 1; Total enrollments in schools in 14 survey blocks of West Godavri district, Andhra Pradesh

Source: 7th All India Educational Survey

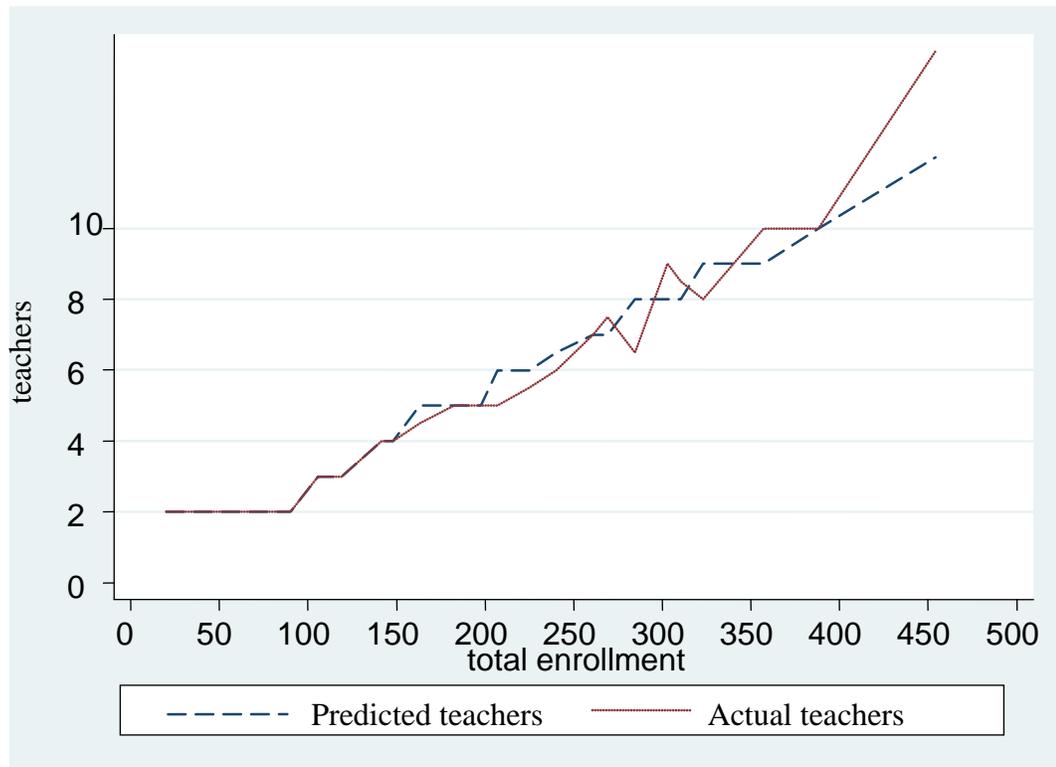


Figure 2: Predicted versus actual teachers as a function of enrollments, census of schools, 14 survey blocks, West Godavri District, Andhra Pradesh

Note: Data are from the 7th All India Educational Survey

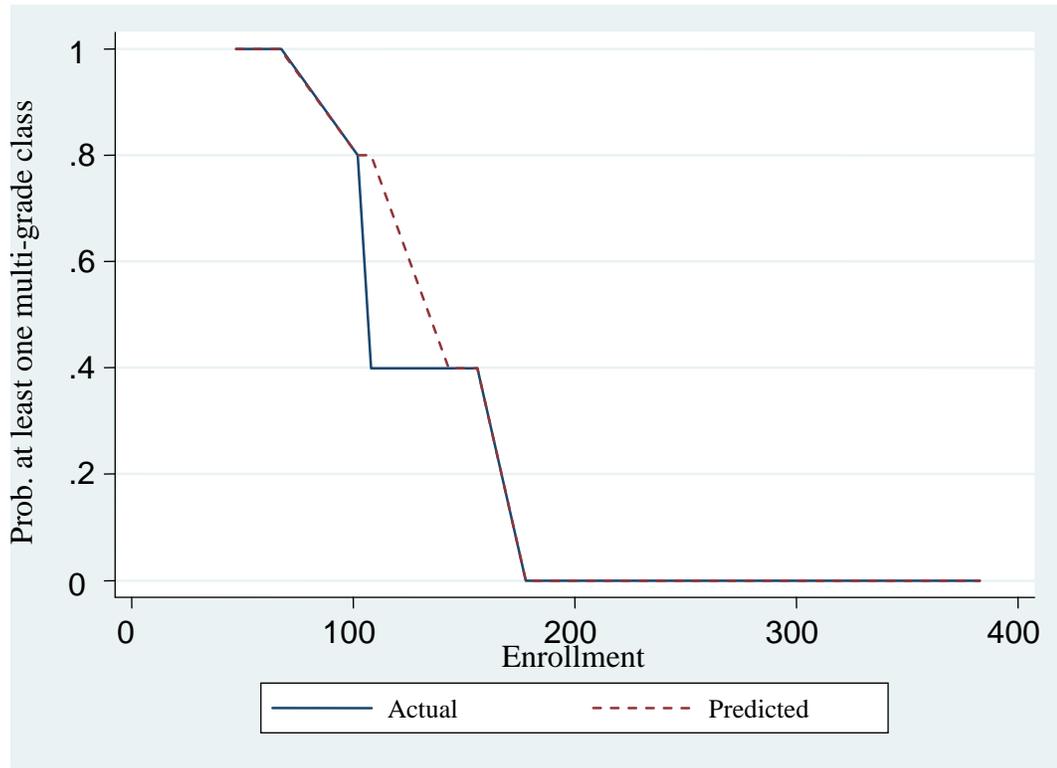


Figure 3: Predicted probability of a multi-grade classroom within a school, versus actual probability, as a function of total school enrollment

References

- Angrist, Joshua D. and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics* May: 533-575.
- Dreze, Jean and Amartya Sen. 1995. *India: Economic Development and Social Opportunity*. Delhi: Oxford University Press.
- Fryer, Roland G., Jr., and Steven D. Levitt. 2005. "The Black-White Test Score Gap Through Third Grade." Cambridge, Mass: National Bureau of Economic Research, Working Paper 11049 (<http://www.nber.org/papers/w11049>).
- Fryer, Roland G., Jr., and Steven D. Levitt. 2004. "Understanding the black-white test score gap in the first two years of school." *Review of Economics and Statistics* 86, no. 2 (May):447-464.
- Hanushek, Eric A. and Steven G. Rivkin. 2006. "School Quality and the Black-White Achievement Gap." Cambridge Mass: National Bureau of Economic Research, Working Paper 12651 (<http://www.nber.org/papers/w12651>).
- Hanushek, Eric A., John F. Kain and Steven G. Rivkin. 2002. "New Evidence About Brown V. Board of Education: The Complex effects of School Racial Composition on Achievement." Cambridge, Mass: National Bureau of Economic Research, working Paper 8741 (<http://www.nber.org/papers/w8741>).
- Heckman, James J. and Richard Robb, Jr. 1985. "Alternative Methods for Evaluating the Impact of Interventions." In James H. Heckman and Burton Singer, eds. *Longitudinal Analysis of Labor Market Data*. New York: Cambridge University Press.
- Hoxby, Caroline M. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper #7867.
- Lavy, Victor and Analia Schlosser. 2007. "Mechanisms and Impact of Gender Peer Effects at School." Manuscript.
- Murphy, Kevin M. and Robert H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business and Economic Statistics* 3(4):370-379.
- Urquiola, Miguel. 2006. "Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia." *The Review of Economics and Statistics* 88(1):171-177.