SCHOOL CHARACTERISTICS AND ACADEMIC ACHIEVEMENT IN PERU: IS THE GEOGRAPHICAL DISTRIBUTION OF RESOURCES REINFORCING SOCIAL EXCLUSION?

MARTÍN VALDIVIA

GIANMARCO LEÓN *

JANUARY 2008

ABSTRACT

This paper goes further in the discussion on the determinants of school attainment in developing countries arguing for the need to take into account the large inequalities in the distribution of traditional school resources and the supply constraints faced by students living in poorer areas. Using a data set containing detailed demographic and socioeconomic characteristics of students, as well as test scores for mathematics and integral communication in Peru, and implementing a two-step correction which accounts for the constraints in school choice, we show that failing to account for these constraints leads to an underestimation of the effect of school resources by about 100%. Not only are the coefficients twice as large but also the contribution of the differences in school resources to the explanation of the differences in math test scores among rich and poor children is doubled. An immediate policy implication is that those hoping to reduce the inequalities in the academic performance of Peruvian children need to consider easing the inequalities in the geographical distribution of traditional school resources. An exclusive focus on school and teacher incentives may help the less poor to improve but at the risk of leaving the poorest behind.

JEL Classification: O10, I21, O54, D13

Key Words: School resources, Academic achievement, Peru

---

* Valdivia: Group of Analysis for Development (GRADE); León: University of California, Berkeley. The first version of this paper was done under the Research Network on Decentralization sponsored by the Economic and Social Research Consortium in Peru (CIES). We are very grateful to participants in seminars given at GRADE, CIES, LACEA and the IADB, who provided insightful and useful comments. Jorge Mesinas provided excellent research assistance during the initial stages of the project. The findings and interpretations contained in this document are those of the authors and do not necessarily represent the views of GRADE or UC Berkeley. The standard disclaimer applies. Correspondence: Valdivia, GRADE, Av. del Ejército # 1870 Lima 27, Perú - PO Box 18-0572. Fax: (511) 264-1882, Tel: (511) 264-1780. E-mail: *jvaldivi@grade.org.pe.*

1) Introduction

Can an increase in traditional school resources play a role in making a difference in the academic achievement of poor children in developing countries? The vast empirical evidence reported in the literature suggest that improvements in traditional school resources (e.g., teacher education and experience, class size, school facilities, among others) have a low chance of effectively helping to improve the academic performance of children in developed and developing countries[1]. This evidence has led the policy debate on education to switch towards the need to work on the structures of school incentives, connecting rewards to teachers or schools to specific outcomes. The appeal of these policies assumes that current resources can often be used more efficiently to improve the performance of children in schools. However, these incentives often result in the exacerbation of inequalities as those that can adjust better to new incentives are the teachers and schools that are already better-off[2].

In this context it is important to reevaluate the evidence on the effect of traditional resources upon the performance of children, especially in developing countries where large inequalities in the distribution of resources across schools – notably between those in poor rural areas and those in large urban areas – make it hard to accept that school and teacher incentives alone can go far in the reduction of inequalities in academic achievement. The lack of evidence of a strong association between school resources and academic achievement does not necessarily imply that they are not important. Rather, what this literature implies is that the relationship between school resources and children's educational attainment is somehow more complicated.

This paper takes a new look at the evidence of the importance of school material and human resources as determinants of school achievement and the associated inequalities

---

[1] See Hanushek (2003) for a vast literature review of the evidence in developed and developing countries. Earlier reviews include Hanushek (1997), Rivkin, Hanushek and Kain (2005) for developed countries, and Hanushek (1995) for developing countries.

[2] Galiani, Gertler and Schargrodsky (2004), for instance, show how the restructuring of incentives associated to decentralization exacerbated inequalities in Argentina.

using data from the 2001 Peruvian evaluation in mathematics and integral communication.[3] We argue that in the context of developing countries -where school infrastructure is still scarce- school and teacher characteristics are not only an important determinant of student performance, but also that the inequalities in the distribution of those resources between different geographic areas pose severe restrictions for poor households to have access to quality education, reproducing the vicious cycles of poverty and exclusion across generations.

Empirically, we estimate the association between traditional school resources and school achievement while controlling for family characteristics. Next, following Card and Krueger (1992) and Contreras (2004), we take into account the selection problem by implementing an identification strategy based on the restriction posed by the geographical distribution of school resources on the school choice. Previous studies have emphasized the fact that students from more educated or richer households, or whose parents are more concerned about their education, are concentrated in higher quality schools, while the opposite would happen with the low quality schools (populated by those with relatively low stocks of material and human resources). However, households vary in their opportunities to choose the quality of schools where they will send their children, as many households reside in areas where there are no high quality schools at all. Failing to take into account this issue will lead to serious biases in the estimation of the school effect on school attainment. To address this problem, we use a two-stage procedure using information on the availability of school resources in the district where the child studies to model the school choice, and use these results to estimate the determinants of educational attainment. Our findings show that not only are school resources important as determinants of school achievement, but also that, when taking into account the geographical constraints to school choice, this effect significantly rises, with important consequences on educational inequality.

[3] These tests evaluate the students in their capacity to solve mathematical problems, logical reasoning, comprehension and communication of mathematical concepts, special reasoning, basic operations, and measurement and estimation, in the case of mathematics. In the Integral communication module, the students are evaluated in their communicational skills and knowledge in Spanish and their native language –if it is the case-; this include, grammar, reading comprehension, analysis of basic sentences, and vocabulary.

The remainder of the paper is organized in six sections, including this introduction. The next section provides a brief literature review on the determinants of educational attainment. Section 3 presents the nature of inequalities in academic achievement and the geographical distribution of traditional school resources. Section 4 describes the datasets that we use and the methodological approach followed, including the identification strategy to tackle the selection problem described above, while Section 5 describes our econometric results. Finally, Section 6 summarizes the results and concludes, providing some policy recommendations.

2) The School and Children's Academic Performance: Literature Review

The relationship between school characteristics and resources and educational quality and learning processes has generated a very rich strand of literature with a great deal of debate within itself. At the empirical level, a significant portion of this literature focuses on the scores of children taking standardized tests, especially in the United States and other developed countries. Hanushek (2003) provides an extensive review of this literature, concluding that there is not a robust association between school resources and educational attainment. More recently, data availability has allowed analysis of this relationship in developing countries but the conclusion remains the same. Nonetheless, this conclusion does not necessarily imply that there are not significant differences between schools or that these differences are not relevant for educational performance. What this literature maintains is that the relationship between the school and children's educational attainment is much more complicated.

Many of the papers summarized by Hanushek (2003) estimate a production function considering family, household, school, and community variables. The authors use school variables such as the average expenditure in the school, the teacher-pupil ratio, teachers' experience and wages, and teachers' level of education. First, in many cases, the selection bias is not properly dealt with. But those who do deal with it tend to be less likely to find a significant and positive association, which is not surprising as the selection bias tends to overestimate the effect of school resources on academic performance. Second, many of the studies analyzed use data aggregated at the state level, which may imply that some of the characteristics observed are not the relevant ones and the relevant ones are not observable by the researcher. Also, it is possible that the differences in the resources are observable, but they are relevant under particular circumstances that are difficult to observe, which in turn allow that some schools find mechanisms through which these differences in resources are relevant.

Another important line of research associated with this discussion argues that it is necessary to make observable what usually had been unobservable to researchers. For example, they look for information about the learning process between teachers and students through variables such as the effective time spent in classes, the use of materials, and homework management, among others. The limitation of this scope is that it is very expensive to collect such information. Also the analysis becomes more complicated if we take into account that the heterogeneity between schools and students could imply that different inputs work differently depending on the context in which they are applied. The effective use of these instruments depends on the resources available in each school, but also on the teachers' motivations. Along these lines, an alternative way of establishing the importance of typically unobserved variables is to analyze the way different monetary and non-monetary incentive structures for teachers and schools affect student performance. The effect of these incentive structures can be estimated without needing to know which specific processes were adjusted, simply through an awareness that the adjustments had most likely been different between schools, depending on the observable and unobservable characteristics of the school, its teachers, and its students.

Nonetheless, the discussion about the relevance of observable resources at the school level of student performance in developed countries does not necessarily apply to developing countries, where the investment in education still falls largely below the average expenditure in OECD countries. Card and Kruger (1992), using longitudinal data from the United States, do find a significant and robust association between school resources and returns to education in the labor market. Among the most important variables considered, we find the teacher-student ratio, teachers' relative wages, and the average duration of the school year.[4] A limitation of this kind of study is that most of them use state-level data, which may hide some important differences between schools within each state. It is important to note that the paper by Card and Kruger analyzes returns to education of the generation of males born between 1920 and 1949, who attended school between 1926 and 1949, a period in which the average level of expenditures in education was lower than and more scattered than the one observed during the seventies and eighties, the period to which most of the studies revised by Hanushek (2003, 1997) correspond. A plausible hypothesis that may reconcile these views is assuming that the relationship between school resources and student performance is non-linear, being more important on the first stages, but insignificant after a certain threshold.[5] This hypothesis becomes important for the analysis of this relationship in developing countries.

Most of the studies reported here do not consider that the type of school where a child attends, and its characteristics, are not strictly exogenous, but rather a consequence of a decision made by parents depending on the availability of resources in the neighborhood. Ignoring the endogeneity of the school the child attends may substantially bias the estimated effect of school characteristics on educational attainment. Moreover, even the direction of the effect may depend on the nature of the selection process. If parents who are more educated, richer, or more concerned about their children's education actually choose better schools, failing to take into account this decision will overestimate the effect of school characteristics. On the other hand, if the decision is constrained by the availability of schools so that families living in poorer areas cannot access schools with better teachers or better infrastructure, then the effect of school characteristics will be underestimated.[6]

Different strategies to identify school effects focus on a specific characteristic or resource, although it often happens that the other school characteristics and resources are

---

[4] See Case and Yogo (1999) for a similar study for South Africa, where they find results similar to those from Card and Kruger (1992).

[5] The STAR study, which applied an experimental design to analyze the effect of class size on the performance of students from a sample of schools in Tennessee, favors the hypothesis of the non-monotonic association between these variables (see Word. et. al., 1990).

[6] Contreras (2004) using a correction by the supply side availability of school resources finds that this correction leads to a higher estimate of the effect of *voucher* schools in Chile.

correlated with the corresponding instrument. Thus, we need to consider an aggregate index that summarizes the information in several characteristics and resources. In the next section, we describe the methodology used to construct the index and analyze the inequalities in the geographical distribution of school resources.

3) The geographical distribution of school resources in Peru

Using the data contained on the National School Census, we start by classifying the schools into three groups using information on school resources that has been proved by previous studies to be relevant proxies for the quality of the school. We control for teacher quality using the percentage of teachers with a university diploma as a proxy. Also, we account for the institutional and administrative quality of the school, and the overall possibilities of transmission of knowledge. For this, we use two variables: if the school belongs to the Intercultural Bilingual Program (EBI, for its name in Spanish), and if it is a one-teacher/multigrade school. The former takes the value of one when the school belongs to EBI program.[7] The one-teacher/multigrade schools have only one classroom and one teacher that serves all grades in the school. Obviously, these schools are the ones with less personnel and fewer resources, which prevents them from having adequate administrative/institutional organization. Moreover, these two types of schools usually exhibit relatively lower school attainment in the national evaluations.[8]

Another variable that should be considered is the physical characteristics of the school itself. For this, we use two variables: the number of computers and the number of libraries available per student. We expect that these variables give us a suitable

[7] This program is designed for schools in rural areas where a majority of the population speaks a native language as their mother tongue. Guerrero (2005) shows that these schools, mainly because of their experimental nature, have very poor school planning, and little availability of qualified teachers.

[8] Similar studies (Hanushek, 2003, among others), mostly regarding developed countries, have taken the size of the class as one of the main school characteristics influencing the school attainment. Nevertheless, we do not include this variable because it is very correlated to the ones that we relate to the quality of the institutional/administrative setting, namely, the ones included in the EBI program and those with just one teacher/one classroom. Also, because in this study we consider that the school choice is an endogenous variable that depends on the school characteristics, class size is assumed to be an outcome variable, not an exogenously determined input.

approximation of the teaching resources and access to information, which have an obvious effect on the quality of the education provided to the students.[9]

In order to have a summary measure of the overall quality of the school, we use a principal component analysis, using the five variables described above. The first principal component obtained explains about 27% of the overall variance of the variables included.[10] This continuous measure of the quality of the school allows us to classify each of the schools in Peru within three categories of quality: high, medium and low. This classification is done based on the empirical distribution of the first principal component, dividing the sample in three quantiles. Figure 1 shows the averages and confidence intervals of the five variables included, by each of the new types of school.

INSERT Figure 1 HERE

About 26% the schools falling into the first category –the low quality schools- work with the EBI program, at the same time, among the high quality ones, there are almost none; the same figure repeats with the type of school: as almost all the schools from medium or high quality are classified as complete,[11] 28% of the low quality ones have only one teacher or have several grades in one classroom. The percentage of teachers who hold a university diploma is certainly lower in the lower quality schools (68%), rising to 87% in the medium quality schools, and around 78% in the high quality ones. What clearly differentiates the top schools from the medium and low quality ones is the stock of learning materials. As the high quality schools on average count 5.5 computers for each 100

[9] It must be noted that we do not reduce the quality of the school resources to these two variables; we just take them as proxies of the quality of the school resources (Glewwe and Jacoby, 1994, show that in Ghana, the addition of libraries in elementary schools significantly increase students´ performance in reading and mathematics; Kingdon, 1996, also shows how the addition of libraries, and computer facilities in India affect educational attainment). For policy purposes, some further research has to be done in order to identify the particular school resources that are more relevant to be improved in particular contexts.

[10] The coefficients associated with each variable after the principal component analysis are shown in Table A. 2 in the appendix.

[11] We call schools complete if they do not fall into the category of one-teacher/multigrade.

students, and 0.60 libraries in the school, the medium and low quality ones have almost none of them.

As described above, there are significant differences between each of the three types of schools in the country. However, this heterogeneity would be neither surprising nor a problem at all if schools were distributed homogeneously along the country. If this were the case, one would expect that the demand for good quality schools would cause the low quality ones to improve their standards. Nevertheless, what we observe is that the geographic distribution of school types is not even close to homogeneous; furthermore, we observe that poorer districts face more severe quality restrictions on school types, while richer districts have the opposite. Figure 2 shows the number of districts with school-choice restrictions by quintiles of unsatisfied basic needs (UBN).[12] The first bar reports the number of districts that only have low quality schools (type 1); while the second bar reports the number of districts that does not have any high quality schools (type 3). Clearly, there is a close association between the wealth of the district and the availability of quality schools within the district. The differences in the geographical allocation of school resources may have an impact on the educational attainment of children living in poorer areas, who are limited to receiving a lower quality of education because they cannot access high quality schools, which in turn reinforces the exclusion mechanisms and the intergenerational transmission of poverty.

INSERT Figure 2 HERE

It must be noted that primary education in Peru is meant to be a public good and elementary education is mandatory by constitutional law. Furthermore, attendance at a particular public school is not restricted to any kind of geographical or political delimitation, as it is in some other countries, such as the US. Along these lines, migration in

[12] The UBN measures the percentage of households living in each district with at least one basic need unsatisfied. This measure has been developed by the INEI, based on the information of the national census of 1993, and periodically updated with new information available.

search of good schools may be considered a problem that may be biasing our results. Nevertheless, evidence from household surveys is consistent with previous qualitative research which shows that geographical migration is mostly due to economic factors, and associated with employment, rather than the search of good quality schools for children. This is more likely to be true for young households, which have children of elementary school age. On the other hand, migration for secondary education is not uncommon, especially in rural areas, where secondary schools are very scarce. Usually, households interested in providing quality education to their children send them to urban areas, or even to the provincial or regional capital to attend school. That implies that school choice at this level is not limited by availability of schools in a particular district. To avoid the biases that may be introduced by migration, our main focus in this paper is on the effect of the school resource availability on educational attainment on primary-school children.

The relative importance of family and school characteristics has captured significant attention in the related literature.[13] Often, simple comparisons of academic performance of children in schools of different quality suggest large school effects but it is already known that the groups are not that easily comparable. On the other hand, simple adjustments for socio-economic status (SES) make this advantage for high quality schools disappear.

We can see such a situation with the Peruvian data, using the information of the 2001 ENRE. Figure 3 shows the estimate of the differences in school attainment of fourth graders by school-type. Simple school differences are large but they disappear when controlling by family SES. In this case, we assess the SES through an asset index (AI), which is constructed using information on asset ownership at the household level.[14] Figure

---

[13] See, for instance, literature reviews in Todd and Wolpin (2004), Rivkin, Hanushek and Kain (2005), among others.

[14] Following Filmer and Prichett (2001), we use a principal components analysis using a based on a set of asset ownership variables. Our variance analysis is based on a polychoric correlation matrix. The polychoric correlation of two ordinal variables is derived as follows. Suppose each of the ordinal variables was obtained by categorizing a normally distributed underlying variable, and those two unobserved variables follow a bivariate normal distribution. Then the (maximum likelihood) estimate of that correlation is the polychoric correlation. If each of the ordinal variables has only two categories, then the correlation between the two variables is referred to as tetrachoric. For further details on the estimation, see Kolenikov and Angeles. (2004). As a robustness check, we also used a simple principal component analysis, and the results remain similar to the ones shown. These results are available upon request of the interested reader.

3 shows great differences in the scores obtained by students across SES quintiles. In both subjects, the difference between the richest and lowest quintiles in the case of schools of type 3, the richer schools, is above 100 points, which is equivalent to 1.5 to 1.7 standard deviations. However, the differences between types of school are smaller within the same SES.[15] Hence, while it is clear that students from poorer schools have lower grades, these gaps are significantly reduced when we control by household SES.

INSERT Figure 3 HERE

Nevertheless, we know from the discussion in section 2 that we cannot conclude anything from the univariate results shown until we control for all the relevant characteristics used in the literature. Following that discussion, in this section we next present the characteristics of the data used and describe the identification strategy followed which is based on the geographical inequalities associated with school resources across districts.

4) Data and Methodological Approach

The previous section suggested the relevance of school resources on children's educational attainment, especially in contexts where there are scarce resources at the school level, as is the case in Peru. This section provides details on the datasets used and the methodological approach to deal with the identification problem.

4.1) Data

The main dataset used in this paper is the National Evaluation of Students' Performance (ENRE) 2001, developed and administered by the Peruvian Ministry of Education (MINEDU), through its Quality Assessment Division (UMC). The main

[15] It must be noted that Figure 3 not only orders the quintiles, but also places them according to the value of the associated score of the first principal component of the AI. This allows us to show that students who attend to the poorer schools basically come from poorer households. Moreover, the wealth level of the richest quintile of households with children in the poorer schools (type 1) is about the same that the one of the poorest quintile of households with children on the richer schools (type 3).

objective of the ENRE is to measure the abilities of Peruvian students in Logic and Mathematics as well as their communicational skills, including reading comprehension and writing. This assessment was done through extensive tests administered throughout the country to students enrolled in fourth and tenth grades in private and public schools[16].

Unlike previous national evaluations, like CRECER and PISA, which used norm based tests, the ENRE uses a criteria model, which not only allows establishing a relative ranking between students, but also was able to assess the extent to which students comply with previously defined standards related to certain areas of an academic curriculum[17]. All of the modules were previously tested in the field and revised by several education specialists, mainly teachers from all over the country. The teachers were also in charge of determining the cut-off points above which it can be said that a student has achieved enough proficiency in each field. Pilot testing allowed improvement in the grading guides, so that they were standardized. Although most of the test consists of multiple-choice questions, there are also some open-ended questions, especially for writing evaluations and oral communication in *Quechua* and *Aymara*. Another important difference between this and previous tests is that the ENRE has been designed taking into account the students' mother tongue. Thus, for fourth graders, the communications section of the test was adapted to native languages in schools where there is a bilingual program. For the logic and mathematics section, the questions were formulated in both languages, so the student could choose whatever language he understands better.

The ENRE also includes questionnaires with basic characteristics of the school, the director, and the mathematics and communication teachers of the evaluated classes. There were also interviews with students' parents conducted.[18] Among teachers´ characteristics, some personal characteristics such as age, gender, mother tongue, educational level, contractual status, attendance in training programs, etc., were included. Also, there are questions about the educational process, for instance, the pedagogic process, the curricular

---

[16] See Torreblanca and Zacarías (2002a).

[17] See Rodríguez and Cueto (2001).

[18] In case the parents are not in the household the survey was applied to the person in charge of the student.

coverage (teaching methods, evaluations, level of use of the materials provided, etc). Regarding parents´ characteristics, we have information on the children's educational history such as age when started attending school, grade repetition, study habits, attitudes towards school and particular fields, etc. Additionally, the survey retrieved socioeconomic and cultural information of the household, like the educational level of the members of the household, occupation, language commonly used in the house, and demographic characteristics of the members. Information about the characteristics of the dwelling was also included: materials of the floor, roof, and walls, and specific asset ownership. Finally, the survey for school directors included questions on social and institutional characteristics of the school (type of school, private/public, participation in the Intercultural and Bilingual Education program (EBI)), about the availability of infrastructure in the classroom and in the school (material of the walls and roof, libraries, toilets, etc.), as well as on the institutional climate in the school (attitudes towards the curriculum, school management, among others).

The sampling design applied for the ENRE was probabilistic, two-staged, clustered and stratified, using as the sampling framework the national census of schools (SISCENS, 2000).[19] The sample size is 10,592 students in fourth grade in 625 schools, which is representative includes of (1) private and public schools; (2) Lima and Callao, big cities, and other cities; (3) multi-teacher public schools, multi-teacher private schools, multigrade and one-teacher schools; (4) Spanish speakers, among the multigrade/one-teacher strata; and (5) Lima and Callao, and other cities within the multi teacher/complete school strata.[20] Descriptive statistics of all the variables used are available in Table A. 1.

Although the full sample of fourth graders is large, only the sub sample that included interviews with the parents is useful for our analysis. This sub-sample includes 5,829 students for the Logic and Mathematics test, and 5,099 for the Integral Communication one.

---

[19] See Torreblanca y Zacarías (2002a and b).

[20] The test was also taken to 10th grade students. It is plausible that our estimation methods will not be as effective to identify the school effect on educational attainment of children enrolled in secondary education, since it is a common practice for students living in Peruvian rural areas to migrate to bigger cities to attend high school. Because of this, we focus the analysis on children enrolled in elementary schools.

In order to be able to retrieve the full distribution of schools, assessing their characteristics and resources available, we also use information from other datasets, such as the school census for the year 2000 and 2002, the height and weight census of 1999, and schools' basic statistics from 2000. Table 1 shows the distribution all operating elementary schools in the country, and compares it with the sample taken for the ENRE 2001. The sample is divided in quintiles of district poverty.[21] The ENRE sample generally reproduces the high concentration of schools on the richest quintile of districts. However, a closer look shows that the ENRE sample is somewhat concentrated in the richest districts. In that group, the participation of the two poorest quintiles on the total is 32%, while that proportion is only 27% in the ENRE sample.

INSERT Table 1 HERE

4.2) Methodology

We estimate a multivariate model trying to disentangle the relative importance of child, household, teacher, school, and district characteristics, on the academic performance of students enrolled in fourth grade. As a proxy of academic performance, we use the test scores for logic/mathematics (LM) and Integral Communication (IC) obtained by the sample of students in the 2001 ENRE sample. The inclusion of school variables on an individual decision model implies some difficulties for the empirical estimation associated with the relevance of some unobserved factors and the selection bias problem.

Let's first deal with the missing variables problem without specific reference to the selection problem. If we understand that the school environment is in fact important, estimating an OLS model will be affected by unobservable characteristics at the household, school, or district levels, generating consistent but not efficient estimates[22]. If the unobserved school characteristics are uncorrelated with the observed household and school

[21] The classification of districts and their distribution in quintiles is based on information about the percentage of households with at least one unmet basic need unmet (UBN) in the 1993 census.

[22] See Greene (2003), chapter 13.4.

variables, then a random effects model will yield minimum variance estimates. Formally, the model to be estimated can be written as follows:

$$r_{ij} = F_{ij}\beta_1 + Z_j\beta_2 + \delta_j + \varepsilon_{ij} \qquad (1)$$

where r_{ij} is the continuous variable associated to the standardized score on the math or communication test of student *i*, who attends to school *j*. F_{ij} is the vector of individual and household observable characteristics, and Z_j is the vector of observable characteristics of school *j*. δ_j denotes the unobserved characteristics of school *j*, which are assumed to be orthogonal to the observed characteristics of the family and school.[23] This model has often been used in the estimation of the education production function[24].

From (1), our interest lies in the magnitude and statistical significance of β_2, which is the vector of coefficients associated with observable school characteristics. For now, we will ignore the endogeneity problem implied in the household decision of the school the child attends. The vector $\hat{\beta}_2$ allows us to determine the sign and statistical significance of the effect of the corresponding variables, although it is not informative regarding the exact magnitude of these effects since its value depends on the units in which Z_j is expressed and their dispersion. Fortunately, there are ways to overcome this problem when we have a continuous dependent variable in a linear model, as is our case. One method consists of analyzing the relative importance of the different variables considered in (1) on the standardized test scores, for example for students from the richest and the poorest quintile, or from the richest and poorest district. Manipulating (1), we are able to decompose the

[23] The necessary assumptions about the error term, ε_{ij}, and the random term δ_j are: $E\left[\varepsilon_{ij} \mid F,Z\right] = E\left[\delta_j \mid F,Z\right] = 0$, $E\left[\varepsilon_{ij}^2 \mid F,Z\right] = \sigma_\varepsilon^2; E\left[\delta_j^2 \mid F,Z\right] = \sigma_\delta^2, \forall i,j,k$: $E\left[\varepsilon_{ij}\delta_k \mid F,Z\right] = 0, \forall i \neq k, j \neq l : E\left[\varepsilon_{ij}\varepsilon_{kl} \mid F,Z\right] = 0$ y $\forall i \neq j : E\left[\delta_j\delta_k \mid F,Z\right] = 0$.

[24] Todd and Wolpin (2004), Rivkin, Hanushek and Kain (2005).

differences in educational attainment between children from the richest quintile (V) and the poorest one (I) as follows[25]:

$$1=\sum_j \frac{\beta_{1j}\left(\bar{F}_{Vj}-\bar{F}_{Ij}\right)}{\left(\bar{r}_V-\bar{r}_I\right)}+\sum_k \frac{\beta_{2k}\left(\bar{Z}_{Vk}-\bar{Z}_{1k}\right)}{\left(\bar{r}_V-\bar{r}_I\right)}+\sum_l \frac{\left(\bar{\delta}_{Vl}-\bar{\delta}_{Il}\right)}{\left(\bar{r}_V-\bar{r}_I\right)}+\frac{\left(\bar{\varepsilon}_V-\bar{\varepsilon}_I\right)}{\left(\bar{r}_V-\bar{r}_I\right)} \qquad (2)$$

where the variables represent the same as in equation (1) and the bars denote averages by quintiles (I and V). Each term on equation (2) represents the particular contribution of each variable (or group of variables) on school attainment of children from the extreme quintiles. Notice that the relative importance of each variable not only depends on β, but also on the relative differences by quintile of each variable. Therefore, the first sum represents the effect of individual and household characteristics, while the second one is the relative importance of observed and unobserved school characteristics. Likewise, solving (2) for each variable included in the analysis, we are able to assess the exact contribution of each variable to the differences between extreme quintiles.

The shortcoming of the econometric estimation of expression (1) is that we do not account for the fact that the household (parents) decide on which school the child should attend, based on their available choice set. Richer or more concerned parents can often decide to send their children to either a public or private school, or choose to send their children to public schools with more observable resources or perceived quality, sometimes traveling daily outside their own neighborhood. Parents with school-aged children can also decide to migrate to another location where the availability of good quality public schools is more widespread. Such selection would make students of different schools intrinsically different so that the estimation of (1) would lead to an overestimation of the effect of school characteristics when we ignore the previous decision stage. Students from more educated or richer households, or whose parents are more concerned about their children's education tend to concentrate in the high-quality schools, while the opposite would happen with the low quality schools (those with relatively low stocks of monetary and human resources).

[25] See Valdivia (2002).

However, recent studies on returns to education, which identify the school decision based on supply constraints, have found even higher returns (see Card, 2001; and Carneiro, et. al., 2003). Initially, these results represented a complex puzzle, but recent interpretations have associated these results with the heterogeneity of the effect of education, which tends to be higher for those groups that have been more affected by supply-side constraints. The underlying idea is that the estimated coefficient using instruments associated with school access and type will correspond to the returns to education of the groups that were more affected by these constraints, and not the average return. Nevertheless, this estimate would arguably be the most useful estimate in evaluating the effects of improvements of school resources for the least favored children.

The studies that have used this supply-based identification strategy have concentrated on the estimation of the returns to education in the labor market, that is, on the effects of the quality of education in the long run. However, Contreras (2004) applies a similar identification strategy to identify the effect of the Chilean school voucher system on the educational performance of high school students. The argument there is that voucher schools are not randomly or homogenously distributed across Chilean localities or regions. In that sense, although some richer and more concerned parents may tend to choose to send their children to voucher schools, not all of them have the same opportunity to choose a voucher school since they are not as available in poorer localities. Contreras finds the estimated effect of *voucher* schools is much higher when adjusting for the heterogeneity in the geographical availability of *voucher* and *non-voucher* schools.

We follow a similar identification strategy to estimate the effect of school resources on Peruvian children's educational attainment using data from the ENRE 2001. The argument here is that public schools do vary in their endowment of material and human resources and in the way parents perceive the quality of the education they provide. However, not all parents face the same choice as better-endowed, quality schools are not randomly or homogenously distributed across Peruvian localities or regions. To follow Contreras, we first need to re-write expression (1), since it includes a lot of school characteristics that have the same endogeneity problem, and to solve the problem for each of them will unnecessarily complicate the analysis. Alternatively, we use the information contained in the School Census to determine the relative wealth of the schools that each student tested on the ENRE attends, and also the wealth of the schools available on the corresponding district. The idea is to instrument the effect of each school using information on the types of schools available on the districts where children from the ENRE sample live, taking into account that there is a different selection process when the district offers only *poor* schools than when it has schools with higher resources. On the first case, given the constraints, there would not be much room for a decision, while in the former it is possible that some unobservable characteristics explain why the family chose to send him/her to a *poorer* school when there were better options in the same district.

Formally, as mentioned above, we form three groups of schools, depending on the school resources and teacher's characteristics. Particularly, we use information from the national school census on the type of school (multigrade, multi-teacher incomplete and complete), whether the school participates in the EBI program (Intercultural and bilingual education), the percentage of teachers with a university diploma, and the availability of computers and libraries in the school. This classification is merged to the schools available in the ENRE sample and is included in a two-stage methodology to compute the effects of the type of school. In the first stage, we estimate an ordered probit model to determine the selection decision of the type of school where each child included in the ENRE attends to following the expression in equation (3):

$$E^*_{ijk} = F_{ijk}\gamma_1 + NE1_k\gamma_2 + NE2_k\gamma_3 + NE3_k\gamma_4 + \mu_{ijk} \qquad (3)$$

where E_{ijk} is the school type of school *j*, where student *i*, resident of district *k*, goes. $NE1_k$, $NE2_k$ and $NE3_k$ represent the number of classrooms in schools of type 1, 2 and 3, operating in district *k*[26]. The inclusion of the number of classes available on each type of school allows us to identify the system and works as a good instrument for our purposes, since it is likely that this variable is related to the school choice made by parents, but it is

[26] We used number of classrooms rather than number of schools as a measure of the availability of schools of different quality in a district because school size was not considered in the principal components analysis. We also tried the estimation using the number of schools as the measure of school availability in districts finding similar results as those reported in section 4. Those results are not reported here for space reasons, but we will be happy to provide them upon request from the interested readers.

plausible to assume that it is orthogonal to students' performance, their ability, or family unobservables. However, the estimation is not done on $E*$, but on E, so we will have:

$$E = 1 \quad if \quad E^* \leq 0$$

$$E = 2 \quad if \quad 0 \leq E^* \leq \lambda$$

$$E = 3 \quad if \quad \lambda \leq E^*$$

Assuming a normal distribution for μ, with mean zero and standard deviation equal to one, we will have the following estimated probabilities:

$$\hat{E}^1 = P\left(E = 1 \mid F, NE2, NE3\right) = \Phi\left(-F_{ijk}\hat{\gamma}_1 - NE2_k\hat{\gamma}_2 - NE3_k\hat{\gamma}_3\right)$$
$$\hat{E}^2 = P\left(E = 2 \mid F, NE2, NE3\right) = \Phi\left(\lambda - F_{ijk}\hat{\gamma}_1 - NE2_k\hat{\gamma}_2 - NE3_k\hat{\gamma}_3\right) - \Phi\left(-F_{ijk}\hat{\gamma}_1 - NE2_k\hat{\gamma}_2 - NE3_k\hat{\gamma}_3\right)$$
$$\hat{E}^3 = P\left(E = 3 \mid F, NE2, NE3\right) = 1 - \Phi\left(\lambda - F_{ijk}\hat{\gamma}_1 - NE2_k\hat{\gamma}_2 - NE3_k\hat{\gamma}_3\right)$$

Using these predicted probabilities, the second stage will use a random effects model with the following characteristics:

$$r_{ij} = F_{ij}\beta_1 + \hat{E}_{ij}^1\beta_{21} + \hat{E}_{ij}^2\beta_{22} + \delta_j + \varepsilon_{ij} \qquad (4)$$

Comparing equation (4) with the one that does not instrument for the school selection will allow us to estimate the relevance of supply side constraints on the academic achievement of Peruvian students. Notice that this identification approach may have some caveats, especially when we consider that there is a chance that the student (alone or with his/her family) might have migrated for educational purposes.

5) School Resources and Academic Achievement: Econometric analysis

5.1) The School effect: Endogeneity controls

In this section we show the results of the estimation of the school effect after applying the two-stage procedure described in equations (3) and (4).[27]

[27] We also estimate regressions for school attainment in IC and mathematics including the regular controls. These regressions yield similar results to the ones shown in previous studies on the subject. For space reasons,

With this classification, we first estimate the selection equation using an ordered probit model in which we include as independent variables the basic individual and family characteristics, and -more importantly- the availability of each type of school in the district (equation 3). The second stage consists in the estimation of equation (4) using the predicted probability for each type of school obtained from the previous stage.

Table 2 reports the coefficients estimated in the first stage regression for the Mathematics and IC tests sample of students enrolled in fourth grade.[28] We first find that family characteristics such as mothers´ schooling, the students´ mothers tongue and the SES indicator appear to be strong determinants of the type of school a child attends. Children of more educated and wealthier parents tend to attend schools of higher quality. Also, children that learn to speak in Quechua or Aymara tend to attend schools of lower quality. We also include in the regression the population density in the district and the percentage of undernourished children living there. The inclusion of this variables responds to political economy concerns about the allocation of schools, since one may assume that the government has the incentive to provide more (and better) schools in more densely populated and wealthier areas because the marginal return of expenditures in these areas, in terms of votes, is much higher. Both of these variables yield statistically significant coefficients with the expected sign.[29]

INSERT Table 2 HERE

Finally, and most importantly, our results confirm that the availability of quality schools in the district, as measured by the number of classrooms available in the district for

---

and in order to not be redundant with the cited papers, we do not include these tables, but we will be more than happy to provide them upon request to the authors.

[28] We must have in mind that we cannot infer magnitudes from the coefficients obtained after an ordered probit estimation, although, the sign and significance of the coefficient give us some clues on the relationship. Table A. 4 shows the marginal coefficients associated with the regressions shown in Table 2.

[29] For consistency purposes, we also run basic OLS and RE regressions using the variables included in previous analysis in Peru. For space reasons, we do not include these tables in the paper, but they are available upon request from interested readers.

each type of school, significantly affects the school the child attends. Our findings show a negative coefficient for the number of classrooms of schools of type 1 in the district meaning it reduces the probability of the child attending a school of type 2 or 3. On the other hand, the presence of classrooms of schools type 3 does raise the probability that a child attends a school of such type. In sum, the distribution of quality schools across the country affects the effective access of Peruvian children to quality education as measured by their material and human resource endowments.

Table 3 shows the results of the regressions associated with equation (4) for both tests, and compares the results obtained using a simple random effects model (RE) and the two stage random effects model described above (RE-IV). The idea of putting these regressions together is to be able to determine the effect of taking into account the family's decision of where to send their children to school, given the supply constraints. There are two possible scenarios here. If the effect estimated when we control for the school choice is lower than the one assuming random distribution of school resources, this would mean that children from more educated or richer parents are the ones who are sent to high quality schools, and thus, the ones who have higher educational attainment. In this case, the family characteristics would be more important in determining the school performance, and the policy implication of such result would be to enhance programs focusing on the household. On the other hand, if the effect after controlling for the endogeneity of the school choice is higher than the one when we disregard it, then the availability of high quality schools would be constraining the children's capabilities of achieving higher scores in standardized tests, and hence the policy implications drawn from these results would point towards a more equitatable distribution of school resources.[30]

INSERT Table 3 HERE

---

[30] See: Card (2001), Kling (2000), and Carneiro, Heckman, and Vytlacil (2001) for further references on the institutional features in the educational systems affecting different population groups.

The results obtained are consistent with the second hypothesis. That is, when we estimate the RE model disregarding the decision about the school, the estimated effects of school characteristics are very small and only significant for the richest type of schools in the Mathematics and Communication tests. We find that the positive effect only reaches 0.32 (0.33) of a standard deviation in the case of the Mathematics (IC) test[31]. The characteristics of the intermediate type of schools do not appear to affect student's performance, when compared to the poorer type of schools.[32] On the other hand, when we use our two-stage method to correct for selection bias associated to the geographical distribution of quality schools, we observe that the two richest types of schools have a significant effect on students' performance, and these effects are much higher than the one estimated when we ignore the selection bias. In fact, the estimated coefficient for schools of type two goes from 0.10 SD to 0.59 SD (0.36 SD), and the effect for schools of type three jumps from 0.32 SD (0.33 SD) to 1.32 SD (1.05 SD), in the Mathematics (IC) test.

As for the family characteristics included in our models, we find that child gender matters, for the case of the Mathematics test. Also, children who learned to speak a native language have lower educational attainments. On the other hand, children from more educated and wealthier parents tend to perform better in school. Nevertheless, when we take into account the endogeneity of the school choice, the estimated coefficients for parents' education and the household asset index drop significantly. Moreover, in the case of the Mathematics test, we do not obtain a statistically significant effect of the household wealth on the educational performance.[33]

This evidence provides empirical support for the hypothesis that geographic inequalities in the distribution of elementary schools with adequate resources is a serious barrier to overcoming the severe inequalities in education in Peru. Moreover, this barrier is

[31] The test scores included as dependent variable in the regression analysis are normalized to have zero mean and variance equal to one.

[32] The omitted category is schools of type 1, which is the one with lowest resources.

[33] For robustness purposes, we also run similar regressions for sub-groups of our sample (the richer, and poorer quintiles, for children of less educated parents, and considering only children attending to two types of schools). These results are consistent with the results presented, and are available upon request.

more important in determining the inequalities in educational outcomes than is family's characteristics for children who live in areas affected by this restriction, such as rural areas or small cities.

As we explained in the previous section, the coefficients estimated in Table 3 are not helpful to establish a relative ranking of the importance of school and child characteristics on students' performance, since it depends not only on the coefficient itself, but also on the other variables that contribute to differences between sub-groups and measurement scales. Nevertheless, computing equation (2), we can report the relative importance of each variable on the educational attainment gap between sub-groups of our sample, such as the one between children from the poorest quintile and the richest one. These estimates are shown in Table 4.

INSERT

Table 4: Contribution of each variable to the inequalities in academic achievement

Table 4 HERE

These results confirm the importance of the identification strategy followed in this paper. First, notice that even after the changes in the specification, the percentage of the overall difference between the poorest and richest quintiles explained by the unobservable characteristics of the school is not very high: about 14% for the Mathematics test, and 6% for the IC one. That is, the omission of the school variables that were not available in the School Census did not significantly increase the estimated contribution of school unobservables.

Second, and most important, notice that the IV correction performed substantially increases the measurement of the contribution of school variables to the differences in academic achievement between the richest and poorest quintiles. For the math test, that contribution increase about 100%, jumping from 21 to 52%. In the case of the IC test, the increase is also very high, going from 18 to 38%. For both tests, the highest contribution comes from best-endowed schools, those of type 3. This increase comes at the expense of the contribution of home inputs, especially of the SES indicator. In other words, failing to control for the constraints that geographic inequalities in the distribution of school resources pose on the school choice by parents tends to underestimate the importance of school variables in explaining differences in academic achievement between poor and better-off children, and overestimates the contribution of home inputs.

6) Summary and Conclusions

This study goes further in the discussion on the determinants of school attainment arguing in favor of the relevance of the availability of traditional school resources. The empirical analysis shows that failing to correct for the school-choice restrictions associated with the geographical inequalities in the distribution of school resources underestimates the effect of school resources on about 100%. Not only are the coefficients twice as large but also the contribution of the differences in school resources to the explanation of the differences in math test scores among rich and poor children doubles from 21% to 52% (18% to 38% in IC).

The literature review provided in the paper illustrates the long discussion on the ways that school characteristics determine the educational quality. The conclusion of the studies regarding developed countries is that school characteristics do not seem to have a significant impact on educational performance. Nevertheless, some of the evidence suggests that, when the distribution of schools was much more unequal and sparse, there was a significant effect. Also, we find that the estimation of the school effect is often biased because of the assumption that the school where the child attends is strictly exogenous. If what happens is that children from more educated or richer parents, or those from parents who are more concerned about the quality of the education actually choose the best schools for their children, the omission of this decision will overestimate the effects of the school resources on student's educational performance. On the other hand, if parents look for the best educational quality for their children, but they are constrained by the availability of schools where they live, this will lead to poorer families living in poor neighborhoods to be limited by the school supply of good teachers and physical resources. This effect will lead to an underestimation of the school effect.

The robust empirical evidence provided in this paper allows us to conclude that there is significant evidence that the school choice within the family is seriously affected by geographical distribution and the constraints on choice that this implies. Even though there are elementary schools in the great majority of districts in Peru, we observe severe differences in the resources they have, such as qualified teachers, school materials, and equipment. Along these lines, we can say that previous estimates of the school effect are underestimating the relevance of teacher's characteristics and school resources, especially on the poorest areas of the country. An immediate policy implication is that the reduction of inequalities in the academic performance of Peruvian children needs to consider reducing the inequalities in the geographical distribution of traditional school resources. An exclusive focus on school and teacher incentives may help the less poor improve but at the risk of leaving the poorest behind.

References

Card, David (2001). "Estimating the return to schooling: Progress on some persistent econometric problems". In *Econometrica* 69 (5): 1127-1160, September.

Card, David; Anna Krueger (1992). "Does school quality matter? Returns to education and the characteristics of public schools in the Unites States". *Journal of Political Economy* 100 (1): 1-40, February.

Carneiro, Pedro; James Heckman; Edward Vytlacil (2003). "*Understanding what instrumental variables estimate: Estimating marginal and average returns to education*". Manuscript, University of Chicago, July.

Case, Anne; Motohiro Yogo (1999). "*Does school quality matter? Returns to education and the characteristics of public schools in South Africa*". NBER Working Paper Series # 7399, October.

Contreras, Dante (2004). "*Vouchers, school choice and access to higher education*". Manuscript, Universidad de Chile, April.

Pritchett, Lant and Deon Filmer (2001). "Estimating Wealth Effects Without Expenditure of Data - Or Tears: An Application to Enrollments in States of India." Demography (February 2001).

Galiani, S; P. Gertler; E. Schargrodsky (2004) "*Helping the Good Get Better, but Leaving the Rest Behind: How Decentralization Affects School Performance*." Preliminary paper presented at LACEA Annual Conference.

Glewwe, Paul and Hanan Jacoby (1994) "Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana," Journal of Human Resources, 29(3): 843-864.

Glewwe, P., & M. Kremer (2006). Schools, teachers, and education outcomes in developing countries. In E. Hanushek, & F. Welch (Eds.), Handbook on the Economics of Education. Amsterdam: Elsevier.

Greene, William H. (2003). *Econometric Analysis*. Prentice Hall, fifth edition, New York.

Guerrero, Gabriela (2005). "An analysis of the implementation of the national policy of bilingual intercultural education in Puno, Peru". Manuscript.

Hanushek, Eric A. (2003). "The Failure of Input-Based Schooling Policies". *Economic Journal* 113: F64-F98, February.

………… (1997) “Assessing the Effects of School Resources on Student Performance: An Update.” *Educational Evaluation and Policy Analysis* 19: 2 (Summer): 141-164.

………… (1995). “Interpreting Recent Research on Schooling in Developing Countries”. In: *World Bank Research Observer* 10 (2): 227–246.

Kingdon, Geeta (1996) “The Quality and Efficiency of Public and Private Schools: A Case Study of Urban India,” Oxford Bulletin of Economics and Statistics, 58(1): 55-80.

Kolenikov, S., and Angeles, G. (2004). “The Use of Discrete Data in Principal Component Analysis With Applications to Socio-Economic Indices”. CPC/MEASURE Working paper No. WP-04-85.

Rivkin, Steven G.; Eric A. Hanushek; John F. Kain (2005). “Teachers, Schools and Academic Achievement”. In: *Econometrica* 73 (2): 417-458, March.

Rodríguez, José; Santiago Cueto. (2001) “¿Cuánto aprenden nuestros estudiantes? Presentación de la Evaluación Nacional del 2001” In Bulletin *Crecer*, Nº2. July.

Todd, Petra E.; Kenneth I. Wolpin (2004). “*The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps*”. University of Pennsylvania, Penn Institute for Economic Research Working Paper 04-019, April.

Torreblanca, A.; D. Zacharías. (2002a) “*Diseño muestral Evaluación Nacional 2001: Ficha técnica-primaria*.” In: http://www.minedu.gob.pe/mediciondelacalidad/2003/

Torreblanca, A.; D. Zacharías. (2002b) “*Diseño muestral Evaluación Nacional 2001: Ficha técnica-secundaria*”. In: http://www.minedu.gob.pe/mediciondelacalidad/2003/

Valdivia, Martin (2002). “*Acerca de La Magnitud de la Inequidad en Salud en el Perú*.” Documento de Trabajo # 37, GRADE, Lima, April.

Word, E.; J. Johnson; H. P. Bain; B. D. Fulton; J. B. Zaharies; M. N. Lintz; C. M. Achilles; J. Folger; C. Breda (1990). “*Student/teacher achievement ratio (STAR), Tennessee’s K-3 class size study: Final summary report 1985-1990*”. Nashville, TN: Tennessee State Department of Education.

Figure 1: Averages and Confidence intervals of variables associated with School Type.

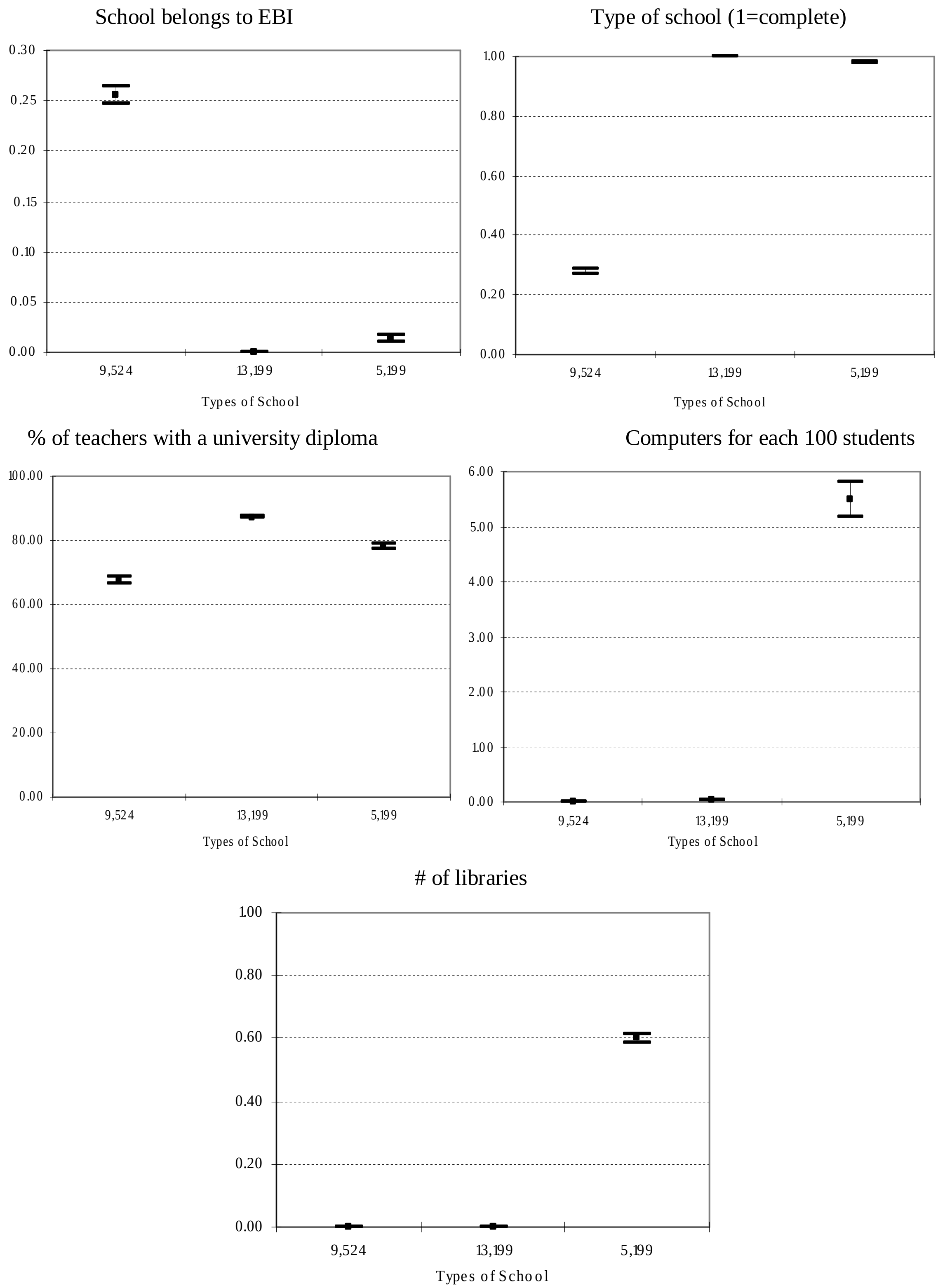


Figure 2: Number of districts with choice constraints, by Quintiles of UBN (population weighted)

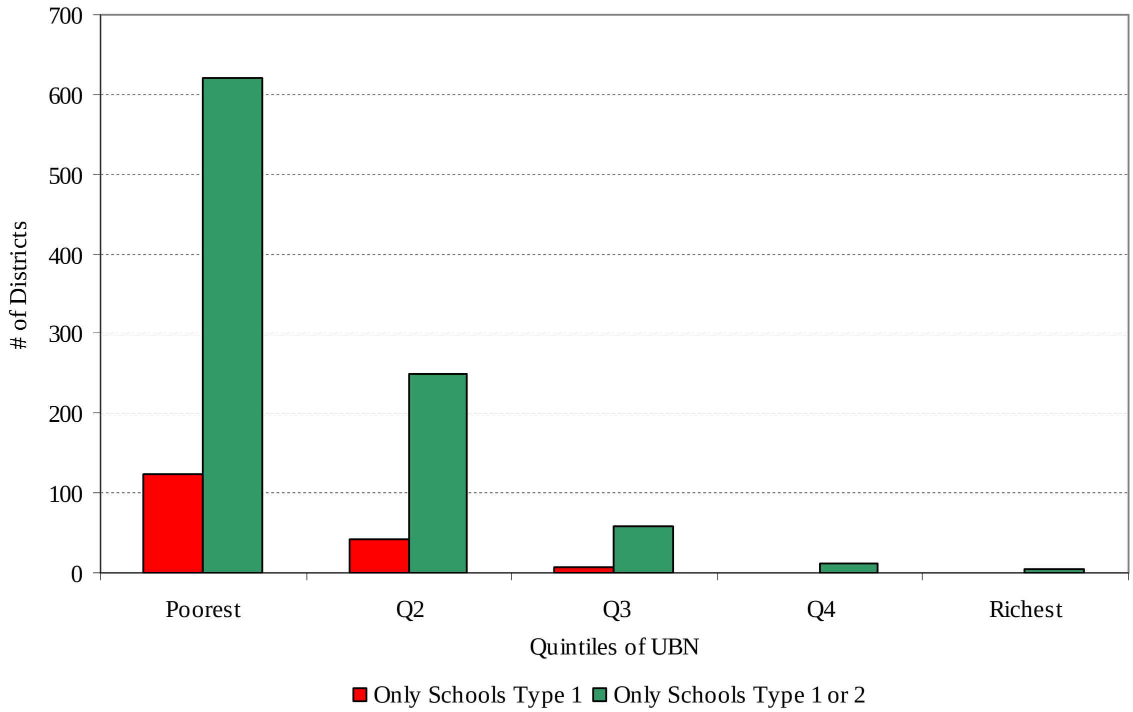
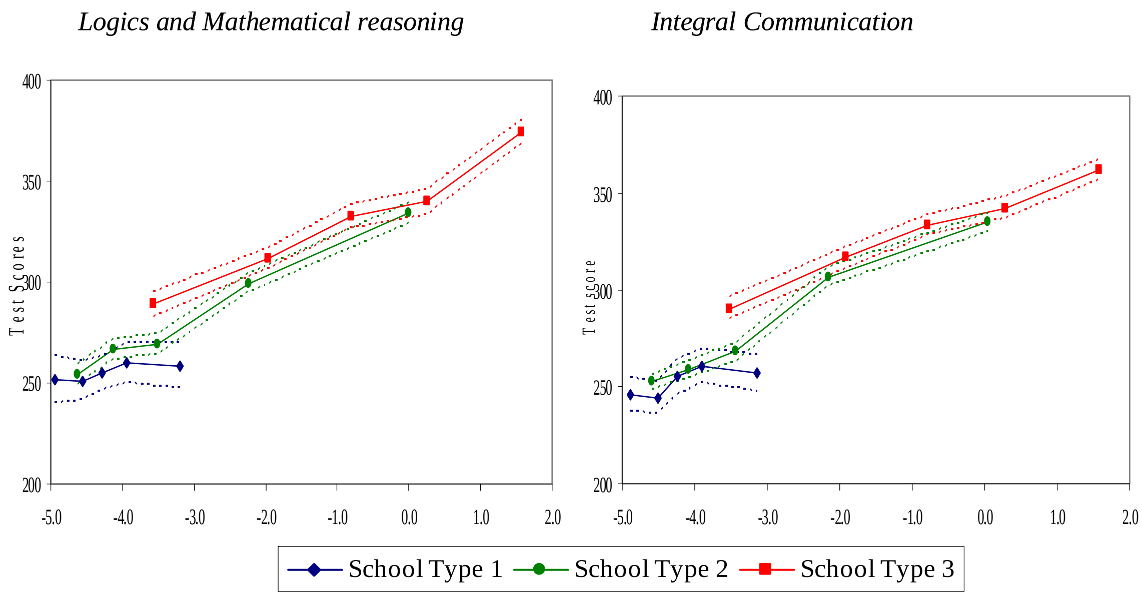


Figure 3: Differences on school attainment in elementary, by school type and SES



Source: Authors elaboration using information from ENRE 2001.

Table 1: Distribution of the number of schools by quintiles of UBN (%) - ENRE 2001 vs. School Census 2002

	ENRE2001			
Quintiles of UBN	Total	(1)	(2)	Census 2002
Poorest	86	13.6	12.9	12.9
Q2	97	15.3	14.6	18.8
Q3	97	15.3	15.3	19.6
Q4	132	20.9	20.1	17.1
Richest	220	34.8	37	31.6
Total	632	100	100	100

Sources: School Census 2002, and ENRE 2001.

(1) Unweighted sample

(2) Weighted sample

Table 2: Ordered probit for the decision of child's school – 4th grade

	Mathematics	IC
Child's Gender (1= boy)	-0.04	-0.03
	(0.96)	(0.63)
Child's mother tongue (1= native)	-0.62	-0.65
	(6.12)***	(6.40)***
Both parents at home	-0.02	-0.03
	(0.43)	(0.53)
Mother's schooling	0.04	0.03
	(1.77)*	(1.33)
Father's schooling	0.03	0.04
	(1.26)	(1.32)
HH asset index	0.20	0.20
	(5.43)***	(5.38)***
# of classrooms in schools type 1 in the district	-0.92	-0.92
(divided by 100)	(3.65)***	(3.65)***
# of classrooms in schools type 2 in the district	-0.15	-0.16
(divided by 100)	(2.42)**	(2.56)**
# of classrooms in schools type 3 in the district	0.15	0.15
(divided by 100)	(3.98)***	(4.06)***
% of urban population in the district	0.01	0.01
	(4.19)***	(4.08)***
% undernourished children in the district	0.01	0.01
	(2.09)**	(2.04)**
Observations	4675	4098
Pseudo R-sq	0.29	0.29
Log likelihood	-3277.04	-2864.17
Chi^2	313.55	307.98

Robust z-statistics in parentheses. Standard errors clustered at the school level. * significant at 10%; ** significant at 5%; *** significant at 1%. Source: ENRE 2001, School census 2002, Basic Statistics 2002.

Table 3: Determinants of school attainment (4th grade)

	Mathematics		IC	
	RE	RE-IV	RE	RE-IV
Child’s Gender (1= boy)	0.14	0.16	-0.03	-0.03
	(7.09)***	(7.68)***	(1.54)	(1.18)
Child’s mother tongue (1= native)	-0.12	0.05	-0.22	-0.09
	(3.82)***	(1.08)	(5.87)***	(1.83)*
Both parents at home	0.00	0.01	-0.04	-0.03
	(0.17)	(0.42)	(1.43)	(1.18)
Mother’s educational level	0.03	0.02	0.03	0.02
	(3.08)***	(1.73)*	(3.26)***	(2.42)**
Father’s educational level	0.02	0.02	0.03	0.02
	(2.57)**	(1.59)	(2.76)***	(2.03)**
HH asset index	0.08	0.02	0.11	0.06
	(6.73)***	(1.16)	(8.82)***	(3.49)***
School type 2	0.10		0.10	
	(1.30)		(1.43)	
School type 3	0.32		0.33	
	(3.66)***		(4.28)***	
Predicted probability school type 2		0.59		0.36
		(3.08)***		(2.24)**
Predicted probability school type 3		1.32		1.05
		(5.66)***		(4.98)***
% undernourished children in the district	-0.01	-0.01	-0.01	-0.01
	(8.34)***	(7.10)***	(8.77)***	(7.91)***
Constant	0.16	-0.59	0.30	-0.21
	(1.49)	(2.79)***	(3.02)***	(1.14)
Observations	4675	4675	4098	4098
Number of schools	524	524	519	519
R sq.	0.34	0.35	0.42	0.42
ρ	0.39	0.39	0.28	0.28
Chi^2	630.78	632.65	1110.87	1092.06

Robust z-statistics in parentheses. Standard errors clustered at the school level. * significant at 10%; ** significant at 5%; *** significant at 1%. Source: ENRE 2001, Weight and Height Census 1999, School census 2002, Basic Statistics 2002.

Table 4: Contribution of each variable to the inequalities in academic achievement

	Mathematics		IC	
	RE	RE-IV	RE	RE-IV
Child, household and environment	*33.86*	*8.60*	*46.00*	*26.27*
Child's Gender (1= boy)	-0.58	-0.63	0.13	0.10
Child's mother tongue (1= native)	3.62	-1.36	5.97	2.42
Both parents at home	0.00	-0.01	0.04	0.03
Mother's educational level	4.85	2.76	5.37	4.13
Father's educational level	3.59	2.28	3.63	2.68
HH asset index	22.37	5.56	30.86	16.90
School	*25.78*	*55.15*	*18.31*	*40.45*
School type 2	-2.66		-2.47	
School type 3	14.52		14.29	
Predicted probability school type 2		-16.47		-9.73
Predicted probability school type 3		58.52		44.08
μ_i (School-level unobservables)	13.92	13.10	6.49	6.10
District	38.64	35.59	33.83	32.41
% undernourished children in the district	38.64	35.59	33.83	32.41
Unexplained	*1.72*	*0.65*	*1.85*	*0.88*
Total	*100.00*	*100.00*	*100.00*	*100.00*

Source: Calculations based on results from Table 4.

Table A. 1: Descriptive statistics, variables used in the regression analysis – Elementary school

	Obs.	Mean	S. D.	Min	Max
Logics and mathematics					
Standardized score	4955	-0.062	0.983	-3.535	3.931
Child's Gender (1= boy)	4955	0.504	0.500	0.000	1.000
Child's mother tongue (1= native)	4955	0.254	0.435	0.000	1.000
Both parents at home	4955	0.820	0.384	0.000	1.000
Mother's educational level	4955	3.264	1.681	1.000	6.000
Father's educational level	4955	3.701	1.642	1.000	6.000
HH asset index	4955	-2.432	2.011	-5.384	2.849
Teacher's mother tongue (1= native)	4955	0.227	0.419	0.000	1.000
Teacher's experience	4955	12.828	7.178	0.000	37.000
Teacher has a title	4955	0.904	0.294	0.000	1.000
Class size	4955	25.842	11.472	2.000	76.000
School type (1= multi-teacher)	4955	0.441	0.497	0.000	1.000
School belongs to EBI	4955	0.177	0.382	0.000	1.000
# of books for each 10 students in the school	4955	0.130	0.403	0.000	5.769
Operative computers for each 1000 students in the school	4955	0.722	2.717	0.000	31.746
% chronically undernourished children in the district	4955	32.030	18.543	1.733	75.119
Integral Communication					
Standardized score	4320	-0.077	0.994	-2.921	3.290
Child's Gender (1= boy)	4320	0.502	0.500	0.000	1.000
Child's mother tongue (1= native)	4320	0.248	0.432	0.000	1.000
Both parents at home	4320	0.818	0.386	0.000	1.000
Mother's educational level	4320	3.289	1.684	1.000	6.000
Father's educational level	4320	3.711	1.642	1.000	6.000
HH asset index	4320	-2.373	2.008	-5.341	2.732
Teacher's mother tongue (1= native)	4320	0.215	0.411	0.000	1.000
Teacher's experience	4320	13.062	7.312	0.000	37.000
Teacher has a title	4320	0.907	0.290	0.000	1.000
Class size	4320	26.208	11.503	2.000	76.000
School type (1= multi-teacher)	4320	0.429	0.495	0.000	1.000
School belongs to EBI	4320	0.172	0.378	0.000	1.000
# of books for each 10 students in the school	4320	0.131	0.400	0.000	5.769
Operative computers for each 1000 students in the school	4320	0.724	2.683	0.000	31.746
% chronically undernourished children in the district	4320	31.909	18.587	1.733	75.119

Table A. 2: Principal component analysis - school type characterization

	Coefficient
School belongs to EBI	-0.35
School type (1=multi-teacher)	0.53
% of teachers with a university diploma in the school	0.12
Operative computers for each 1000 students	0.47
# of libraries in the school	0.6
% of the overall variance explained by the first PC	26.9

Table A. 3: Asset index principal component analysis

Variable		LM Sample	IC sample
Floor	Dirt	-0.295	-0.293
	Rough wood	-0.132	-0.129
	Cement	-0.069	-0.065
	Loseta	-0.015	-0.011
	Vinyl	-0.006	-0.002
	Parquet	0.194	0.196
Walls	Estera	-0.518	-0.512
	Eternit	-0.438	-0.430
	Wood	-0.355	-0.348
	Stones and mud	-0.288	-0.283
	Quincha	-0.255	-0.252
	Adobe	-0.151	-0.148
	Cement or concrete	0.108	0.109
Roof	Straw	-0.383	-0.375
	Estera	-0.284	-0.278
	Broad	-0.271	-0.265
	Cane	-0.261	-0.254
	Calamina	-0.174	-0.168
	Roofing tile	-0.084	-0.079
	Wood	-0.054	-0.049
	Cement or concrete	0.125	0.125
Water supply	River	-0.321	-0.318
	Bought from a truck	-0.228	-0.226
	Water well	-0.200	-0.197
	Water well inside the house	-0.163	-0.159
	Public network outside the house	-0.131	-0.127
	Public network within the house	0.080	0.081
Connected to public dwelling	No	-0.285	-0.282
	Yes	0.120	0.121
Light	Candle	-0.475	-0.475
	Kerosene lamp	-0.268	-0.267
	Gas lamp	-0.188	-0.187
	Battery	-0.185	-0.184
	Electricity	0.089	0.089
Car	No	-0.186	-0.180
	Yes	0.160	0.159
Bicycle	No	-0.213	-0.208
	Yes	0.098	0.098
Kitchen	Doesn't have	-0.500	-0.507
	With wood	-0.259	-0.263
	With Kerosene	-0.133	-0.134
	With gas	0.104	0.107
Truck	No	-0.074	-0.071
	Yes	0.076	0.074
PC	No	-0.193	-0.189

	Yes	0.179	0.181
CD player	No	-0.246	-0.247
	Yes	0.127	0.130
Washing machine	No	-0.183	-0.179
	Yes	0.160	0.161
Iron	No	-0.307	-0.308
	Yes	0.122	0.123
Radio	No	-0.201	-0.204
	Yes	0.021	0.022
Refrigerator	No	-0.258	-0.259
	Yes	0.152	0.156
TV	Doesn't have	-0.356	-0.354
	Black and White	-0.163	-0.160
	Color	0.135	0.137
Telephone	No	-0.225	-0.223
	Yes	0.177	0.179
Cellular phone	No	-0.189	-0.185
	Yes	0.165	0.166
VCR	No	-0.218	-0.215
	Yes	0.167	0.169
Works in agriculture	No	0.291	0.290
	Yes	-0.097	-0.098

Table A. 4: Marginal coefficients after ordered probit

	Type of school					
		Mathematics			Integral Communication	
	Pr(E=1 | X)	Pr(E=2 | X)	Pr(E=3 | X)	Pr(E=1 | X)	Pr(E=2 | X)	Pr(E=3 | X)
Child's Gender (1= boy)	0.005	0.009	-0.014	0.003	0.007	-0.010
	(0.95)	(0.95)	(-0.96)	(0.63)	(0.63)	(-0.63)
Child's mother tongue (1= native)	0.085	0.097	-0.182	0.088	0.104	-0.192
	(4.25)***	(5.88)***	(-6.75)***	(4.29)***	(6.07)***	(-7.04)***
Both parents at home	0.003	0.005	-0.008	0.003	0.007	-0.011
	(0.43)	(0.42)	(-0.42)	(0.54)	(0.52)	(-0.53)
Mother's schooling	-0.005	-0.010	0.014	-0.004	-0.008	0.011
	(-1.73)*	(-1.74)*	(1.76)*	(-1.31)	(-1.31)	(1.32)
Father's schooling	-0.004	-0.008	0.011	-0.004	-0.008	0.012
	(-1.25)	(-1.24)	(1.25)	(-1.31)	(-1.30)	(1.31)
HH asset index	-0.021	-0.043	0.064	-0.021	-0.046	0.066
	(-4.54)***	(-4.64)***	(5.20)***	(-4.43)***	(-4.67)***	(5.17)***
# of classrooms in schools type 1 in the district (divided by 100)	0.099	0.203	-0.302	0.095	0.209	-0.304
	(3.37)***	(3.58)***	(-3.75)***	(3.35)***	(3.58)***	(-3.74)***
# of classrooms in schools type 2 in the district (divided by 100)	0.017	0.034	-0.051	0.017	0.037	-0.054
	(2.38)**	(2.28)**	(-2.38)**	(2.50)**	(2.43)**	(-2.53)**
# of classrooms in schools type 3 in the district (divided by 100)	-0.016	-0.032	0.048	-0.015	-0.033	0.048
	(-3.79)***	(-3.57)***	(3.91)***	(-3.79)***	(-3.68)***	(4.00)***
% of urban population in the district	-0.001	-0.002	0.004	-0.001	-0.002	0.004
	(-3.95)***	(-3.74)***	(4.13)***	(-3.92)***	(-3.68)***	(4.04)***
% undernourished children in the district	-0.001	-0.002	0.003	-0.001	-0.002	0.003
	(-2.02)**	(-2.03)**	(2.07)**	(-1.98)**	(-1.99)**	(2.03)**
Observations	4675	4675	4675	4098	4098	4098
Pseudo R-sq	0.29	0.29	0.29	0.29	0.29	0.29
Log likelihood	-3277.04	-3277.04	-3277.04	-2864.17	-2864.17	-2864.17
Chi^2	313.55	313.55	313.55	307.98	307.98	307.98

Robust z-statistics in parentheses. Standard errors clustered at the school level. * significant at 10%; ** significant at 5%; *** significant at 1%
Source: Calculations based on results from Table 2