INEQUALITY IN SCHOOL RESOURCES AND ACADEMIC ACHIEVEMENT: EVIDENCE FROM PERU*

GIANMARCO LEÓN
UNIVERSITAT POMPEU FABRA AND BARCELONA GSE

MARTÍN VALDIVIA
GRUPO DE ANÁLISIS PARA EL DESARROLLO (GRADE)

ABSTRACT

This paper goes further in the discussion on the determinants of school attainment in developing countries. To properly estimate the effects of school resources on academic achievement, we the need to take into account the large geographical inequalities in the distribution of school resources and the supply constraints faced by students living in poorer areas. We do so by implementing a two-step correction that accounts for the constraints in school choice. Our findings suggest that failing to account for these constraints leads to an underestimation of the effect of school resources on school achievement of about 100%. This underestimation is particularly important for girls and in Math. Additionally, the contribution of school resources in explaining the gap in test scores between rich and poor students are doubled once we account for the constrained choices.

JEL Classification: O10, I21, O54, D13
Key Words: School resources, School choice, Academic achievement, Peru

---

* León: Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE, E-mail: gianmarco.leon@upf.edu; Valdivia: Grupo de Análisis para el Desarrollo (GRADE), E-mail: jvaldivi@grade.org.pe. 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, UC Berkeley, CIES, LACEA XII, MWIEDC, and the IADB, who provided useful and insightful comments. León thanks the financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075) and grant ECO2011-25272. Jorge Mesinas provided excellent research assistance during the initial stages of the project. The standard disclaimer applies.

1) Introduction

Despite the large body of studies assessing the effect of traditional school resources on academic achievement, there is still an active debate on whether they play a significant role in improving academic achievement of poor children in developing countries. Most of the available evidence suggests that improvements in traditional school resources (e.g., teacher education and experience, school facilities, etc.) have a low chance of effectively helping improve the academic performance of children in developed and developing countries.[1] This evidence has led the policy debate to lean towards the need to work on the structures of school incentives, connecting rewards to teachers or schools to specific outcomes (e.g. Duflo et al 2014, Das et al 2013). However, these incentives can result in the exacerbation of within and between school inequalities, since those who often adjust to the new incentives are already better off students, schools and teachers.[2]

In this paper, we take a new look at the evidence on the importance of school material and human resources as determinants of school achievement and the associated inequalities. Using data from public schools in contained in the Peruvian school census, along with pupil-level characteristics and standardized test scores, we show that school and teacher characteristics are important determinants of student performance. However, these characteristics become empirically relevant only once we properly account for the constraints in school choices faced by parents. Using a two-stage procedure, we model a constrained school choice and estimate the determinants of educational attainment. Our results show that failing to account for these constraints leads to an underestimation of the effect of school resources on school achievement of about 100%. This underestimation is particularly important for girls, and in Math. Furthermore, the contribution of school

---

[1] Hanushek (2003) provides a very detailed literature review of the evidence in developed and developing countries. Earlier reviews include Hanushek (1997), Rivkin, Hanushek and Kain (2005), Case and Yogo (1999) 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. Glewwe et al (2009) show that providing textbooks to school children does not affect average test scores, but increase within school inequalities.

resources in explaining the gap in test scores between rich and poor students doubles once we account for the geographical distribution of resources.

The relationship between school characteristics and educational quality has generated a very rich strand of literature with a great deal of debate about the interpretation of the empirical results. The Coleman Report (Coleman et al 1966) found that family characteristics are more relevant determinants of academic achievement than school resources. In a reevaluation of the evidence 40 years later, Gamoran and Long (2006) and Gamoran (2001) find that these results are still relevant, and predict that the pattern will hold under different forecasting scenarios. However, Hanushek (2003) provides an extensive review of the evidence for developed countries (not only the US), concluding that the association between school resources and educational attainment is not robust enough to draw conclusions. Nonetheless, this does not necessarily imply that there aren't significant differences between schools or that these differences are not relevant for educational performance. The analysis of this relationship in developing countries is mixed, with findings showing that certain types of traditional school resources make a difference in educational achievement, while others are not relevant at all. Kremer et al (2013) reviews a large number of recent randomized control trails, and conclude that traditional school resources do not increase test scores, mainly because students are already lagging behind by the time the interventions are implemented. Moreover, some of these interventions increase within-classroom inequalities, benefiting only already better-off students.[3] On the other hand, Glewwe et al (2011) in a review of the literature for developing countries finds that, in some contexts, school infrastructure or teacher characteristics have a positive and significant effect on academic achievement.

Many of the papers reviewed by Hanushek (2003) estimate a production function for educational attainment, as measured by the scores children obtain in standardized tests, considering family, household, school, and community variables, using either a

---

[3] Many studies argue that increases in educational resources have a limited impact on learning in distorted educational systems (Hanushek 1995 and Pritchett and Filmer 1999).

contemporaneous or value-added specification.[4] Authors measure school quality with variables such as average public and/or private expenditure in the school, teacher-pupil ratio, and teachers' formal training, experience and wages. The key difficulty in this estimation is that the type of school where a child attends, and its characteristics, are not exogenous, but it is result of a decision (by their parents), which makes it challenging to disentangle the effect of parent's preferences from the school quality. Ignoring this decision stage in the estimation may substantially bias the effects of school characteristics on educational attainment.

Further, the sign of that bias is not clear, since depends on the nature of the selection process. On one hand, students from more educated or richer households, or whose parents put a higher value on their education, are concentrated in higher quality schools, while the opposite would happen with the low quality schools. If this is the case, ignoring the school choice decision in the estimation will lead to an overestimation the effect of school characteristics, attributing the effect of family background to school resources. On the other hand, if the decision is constrained by the availability of schools, so families living in poorer areas cannot access schools with better teachers or better infrastructure, regardless of their preferences, then the effect of school characteristics will be underestimated.

Most of the literature in education and economics so far has only accounted for the demand side of the selection, for example analyzing the effects of conditional cash transfers on school performance (Behrman, Parker and Todd, 2011). The studies that exploit a demand side shock to identify the school selection tend to find larger estimates of school resources on academic achievement. Studies that exploit supply side shocks to identify the school resources often rely on aggregated data (Hanusheck, 2003), which may hide relevant relationships.

The discussion about the relevance of observable resources at the school level of student performance in developed countries can potentially be quite different in developing countries, where the investment in education still falls largely below the average

[4] Todd and Wolpin (2007) provide a detailed description of each of these approaches and the assumptions required for the corresponding estimates to be reliable.

expenditure in OECD countries (Glewwe and Kremer, 2006). Studies in developing countries with clear identification strategies based on randomized trials offer mixed results. For instance, provision of textbooks or workbooks improved children's academic performance in Nicaragua (Jamison et al 1981), Philippines (Tan et al 1997), but not in Kenya (Glewwee et al 2009). Radio instruction in Nicaragua (Jamison et al 1981) and computer-assisted learning programs in India (Banerjee et al 2007) showed the important contribution technology could make to improve learning in developing countries. On the other hand, experiments in Kenya showed little impact on test scores from reductions in class size (Duflo et al 2014), flip charts (Glewwe et al 2004) and deworming medicine (Kremer and Miguel 2004).

Recent studies on returns to education, using an instrumental variable approach, and identifying the school decision based on supply constraints, have found that the IV results are larger than the OLS (see Duflo 2001, Card, 2001; and Carneiro, et. al., 2003). Card and Krueger (1992), using longitudinal data from the United States, find a significant and robust association between school resources and returns to education. Initially, these results represented a complex puzzle, however, one can interpret them as being associated with the heterogeneity of the effect of education and the decreasing returns to educational resources. Particularly, one would expect the effect of supply side constraints to be larger for 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 quality 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 one in evaluating the effects of improvements of school resources for the least favored children.[5]

Heyneman and Loxley (1982, 1983a and 1983b), using cross-country data, have argued that in low-income countries the effect of school and teacher quality on academic

[5] Card and Kruger (1992) analyze the 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.

achievement in primary school is comparatively greater. These seminal studies have been followed by a great amount of work studying the "Heyneman-Loxley Effect", finding supporting evidence for this hypothesis using both cross-country and within-country evidence.[6]

A hypothesis that might reconcile these opposing views is that the relationship between school resources and student performance is non-linear, being more important at lower levels, but insignificant after a certain threshold.[7] This hypothesis becomes important when interpreting the current evidence from developed countries and analyzing this relationship, and trying to design policies in developing countries.

The studies using supply-side instruments to identify the effects of school resources have concentrated on the estimation of the returns to education on the labor market, that is, on the effects of the quality of education in the long run. Contreras (2004) applies a two-step correction that accounts for the constrained supply of schools in Chile 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

Different strategies to identify school effects focus on a specific characteristic or resource, although it often happens that other school characteristics and resources are correlated with the corresponding instrument. Using detailed school census data, matched

---

[6] See for example: Fuller and Heyneman 1989, Baker and LeTendre 2002, Baker et al 2001, Chudgar and Luschei 2009 and 2011.

[7] 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). This hypothesis is also suggested in Hanushek 2003.

with pupil level test scores, we overcome the selection problem discussed above by using an index of school resources aggregated at the district level to identify the effect of school resources on academic achievement. An advantage of our identification strategy over the previous literature is that resources for public schools at the district level are defined by the different instances of the Ministry of Education, and respond to past political economy and demographic characteristics of the district, rather than current school performance, which allows us to separate the effect of parental preferences and characteristics from school resources.

The paper is organized in six sections, including this introduction. The next section describes the datasets used in the empirical analysis, while Section 3 provides a background on the Peruvian educational system, and discusses the inequalities in academic achievement and geographical distribution of traditional school resources. Section 4 describes the methodological approach to tackle the selection problem and the empirical model. Section 5 discusses our findings, and finally Section 6 summarizes the results and concludes, providing some policy recommendations.

2) Data and Descriptive Statistics

In our main analysis, we combine detailed, pupil level information and test scores collected in the 2001 Evaluación Nacional de Rendimiento Educativo (ENRE), combined with the 2002 National School Census.

Unlike previous national evaluations, like CRECER,[8] which used norm based tests, the ENRE uses a criteria model, which not only allows establishing a relative ranking between students, but is also able to assess the extent to which students comply with pre defined standards related to areas of the academic curriculum.[9] All modules of the test were pre-tested in the field and revised by several education specialists. Teachers were in charge of determining the cut-off points above which a student can be said to have achieved

[8] CRECER is a national evaluation applied by the ministry of education to a sample of primary school children in Peru. This test was used between 1996 and 1998.

[9] See Rodríguez and Cueto (2001).

enough proficiency in each field. 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 language section of the test was adapted to native languages in schools with a bilingual program. For the logic and mathematics section, the questions were formulated in both languages, so students could choose the language she understood better.

The ENRE also included a household questionnaire, applied to parents or guardians. It includes information on the child's educational history such as age when started school, grade repetition, study habits, attitudes towards school and particular fields, etc. Additionally, the survey collected information on the household's socioeconomic status and characteristics of the dwelling (floor materials, roof, and walls, specific asset ownership, etc.), and questions related to the characteristics of other household members: educational level, occupations, language, and demographic characteristics. This household questionnaire was applied to a random subsample of the population of students who took the academic tests.

The sampling of the ENRE is probabilistic, two-staged, clustered and stratified, using as sampling framework the previous national school census (SISCENS 2000).[10] Overall, the ENRE has information for 10,592 fourth graders from 625 schools. The survey is representative 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.[11] Table 1 shows the detailed definitions of all the variables used in the empirical analysis, while in

---

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

[11] 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.

Table A. 1 we provide the descriptive statistics of these variables. Although the full sample of fourth graders is large, only a random sub sample included interviews with the parents. This sub-sample includes 5,829 students with Math test scores, and 5,099 with the Language test scores, and is the main analysis sample that we will use in the remainder of the paper.[12]

Even though the ENRE is representative at the national and regional level, this does not ensure that the sample resembles the actual distribution of schools throughout the country. Using information from the 2002 National School Census, collected by the MINEDU, Table 2 shows that operating public elementary schools in the (weighted) ENRE sample are similarly distributed as the ones in the census. In this table, we divide the sample into quintiles of district poverty.[13] The ENRE sample generally reproduces the high concentration of schools on the richest quintile of districts. In that group, the participation of the two poorest quintiles on the total is 32 percent, while that proportion is only 27 percent in the ENRE sample.

The statistics unit of the Peruvian Ministry of Education has been aplying the School census since 1993, becoming an annual effort since 1998. It is based in the reports provided by school directors by june of each year, and includes information on enrollment, teachers, administrative personnel, infrastructure and furniture. The infrastructure information includes the number and condition of classrooms, libraries, laboratories, sports facilities, offices, toilets, etc. The 2002 school census covered 95% of the schools of all levels in Peru.

Finally, in order to get measures of the socioeconomic conditions in each district, we use information from the 1993 Population Census.

---

[12] The number of observations is different from the one effectively used in the empirical analysis because a handful of these observations have missing values in some variables of interest.

[13] 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 population census.

3) Background and the Geographical Distribution of School Resources in Peru

Peru is a middle-income country with relatively low levels of academic achievement. The average adult in the country has completed 7.7 years of formal education, compared to 8.2 years in the average Latin American country (UNDP 2001). Further, the quality of education is far below the one in countries with similar per capita income. In the latest PISA evaluation, Peru ranked last, below poorer countries, like Indonesia or Jordan.

Primary education in Peru is a public good; attendance is mandatory (for primary school), and not restricted to a geographical or political delimitation, as it is in some other countries, such as the US. Overall, there are 34,337 schools throughout the country (School Census, 2002). Given that 82.8% of the schools in the country are public, and serve to 86.7% of the pupils, in this paper we focus our attention on these schools. Hence, policies implemented over public school are the ones that are going to impact the majority of school children in the country. Further, private schools are not widespread in the country, and especially in rural areas. Even though there are 5,888 private schools, most of them are located in urban centers (62%), and large cities. Andrabi et al (2008) and Pal (2010) show that the location of private schools is not random at all, but rather they respond to local conditions that allow them to reach distant localities (eg. highways and roads), and that has the necessary inputs to provide a relatively higher quality service at a lower cost (eg. low cost teachers). In the Peruvian case, transportation infrastructure in the Andes is very poor, and there are no constraints for women to work, hence the conditions described in the papers cited above can't be reproduced in Peru, which is one of the reasons why private schools aren't widespread.

The quality of public schools has significant variation, and classifying them according to the availability of resources is challenging. There are a number of resources that we can potentially consider, and generating a simple ranking of schools requires us to be able to determine the relative importance of each of the resources considered on the *"quality"* of the school. In this section we use a principal component analysis to construct a uni-dimensional school quality index, and explain which are the variables considered.

The National School Census contains detailed information on all public and private schools in the country, and we use this data to generate an index that will allow us to rank schools, and classify each school in the ENRE into one of three categories. This index is based on school characteristics that have been shown in previous studies to affect academic achievement. The school characteristics that we consider fall into one of three categories: (i) Teacher quality, measured by the percentage of teachers with a university diploma;[14] (ii) Institutional and administrative quality, which we proxy by an indicator for whether the school is complete or one-teacher/multigrade school;[15] (iii) School resources, proxied by the number of computers per 1,000 students and the number of libraries available per 1,000 students.[16]

In order to have a summary measure of the overall quality of the school, we use the first principal component from the four variables described above. The first principal component explains 28.2% of the overall variance of the variables included. The coefficients associated with each of the four variables included to in the principal component analysis are shown in Table A. 2 in the appendix. Figure **1** shows the kernel density of the first principal component. It is clear from the figure that there are three points

[14] The literature documenting the effects of teacher qualifications and credentials on academic achievement is fairly equivocal. However, recent studies are tilting the debate towards showing that there is an association between teacher quality and academic performance. For instance, Barnett (2003) argues that more qualified pre-school teachers improve student performance in the future. Buddin and Zamarro (2009) use longitudinal data from Los Angeles County, and uses a value added approach to show that the relationship between academic achievement and teacher licensing is weak, but experience is positively associated with performance. Clotfelter et al (2010) use data from high schools in North Carolina and find compelling evidence that teacher credentials, particularly licensure and certification, affects student achievement in systematic ways and that the magnitudes are large enough to be policy relevant.

[15] Guerrero (2010) shows that single-teacher schools in Peru usually have a very poor administrative quality. Several studies (Hanushek 2003, Angrist and Lavy 1999), have shown that the size of the class is one of the main school characteristics influencing school attainment. Nevertheless, we do not include this variable because of it may capture other effects, such as being in a rural area. The variable of multigrade classrooms is more accurate for measuring the teaching possibilities in the classroom. Also, since 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.

[16] We use these variables as sufficient statistics for 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. Banerjee et al (2007) show that the availability of computers in rural Indian schools help improve test scores.

with particularly high probability mass (and very similar scores), which suggests that there is a natural classification of schools into three categories. Hence, we split the universe of public schools in the country into three groups: high quality (3,807 schools), medium quality (13,935 schools), and low quality (10,180 schools).[17]

There are noticeable differences between school categories. Figure **2** shows the average and confidence intervals of the four variables considered into our index, for each of the school types defined. High quality schools score very high in all of the areas: all schools are complete, 85.6% of teachers have a university diploma, they have 6.2 computers and 9.1 libraries per 100 students enrolled. At the other end, only 32% of low quality schools have more than one teacher, 61% of these teachers have a university diploma, and there are 0.2 and 0.4 computers and libraries per each 100 students, respectively. Schools in the medium quality category have high teaching potential, and all of them have one or more teachers per classroom, but the teaching infrastructure is still scarce.

The geographic distribution of school quality is also far from homogeneous. Poorer districts face more severe quality restrictions, while richer districts have availability all types of schools. Panel A in Figure **3** shows the number of schools of each of the three types, per quintiles of district wealth (based on the percentage of households with at least one unsatisfied basic need - UBN);[18] while in Panel B we plot the number districts in each quintile with school-choice restrictions. The first bar reports the number of districts that only have low quality schools (type 1), while the second one reports the number of districts that do 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 distribution of school resources may have an impact on the educational attainment of children living in poorer areas, who are constrained to receiving a lower quality of education because of the unavailability of high quality schools.

[17] We use terciles to divide the sample intro three groups. Given the accumulation of mass in three points (as its clear from Figure **1**), and particularly in the middle of the distribution, the terciles do not have the same number of observations.

[18] 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 National Statistical Institute (INEI), based on the information of the 1993 population census, which has been updated with new information available.

One concern with our argument is that, even though children in poorer districts face a quality constraint, their families can overcome this constraint by migrating to areas where better schools are available. Nevertheless, census evidence is consistent with previous qualitative research, showing that geographical migration is mostly due to economic factors, associated with employment, rather than the search of good quality schools for children (Yamada, 2010). On the other hand, migration for secondary education is not uncommon, especially in rural areas, where secondary schools are very scarce or not existent (compared to primary schools, which are available in almost every district). Usually, households interested in providing higher education to their children send them to urban areas, or even to the provincial or regional capital to attend secondary school. To avoid the biases that may be introduced by migration decisions, we focus in this paper 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 related literature.[19] 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. Further, McEwan et. al., 2008 find that simple adjustments for socio-economic status (SES) make the advantage of high quality schools disappear.

We can see a similar result with Peruvian data. Using information from the 2001 National Evaluation of Students' Performance (ENRE). Figure **4** shows the estimate of the differences in school attainment of fourth graders by school type and socioeconomic status (SES), as measured by an asset index (AI).[20] Among children studying in high quality

---

[19] See, for instance, literature reviews in Todd and Wolpin (2007), Rivkin, Hanushek and Kain (2005), among others.

[20] Following Prichett and Filmer (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

schools (type-3), the difference between the richest and the poorest quintiles is above 70 points, which is equivalent to 1.5 to 1.7 standard deviations. However, the differences between types of schools are smaller within the same SES.[21] This is, for any given level of the SES, there is statistical difference in test scores between children who go to high or low quality schools. Hence, while it is clear that students from poorer schools have lower grades, these gaps are significantly reduced when we control by household SES.

Nevertheless, we know from our discussion above that we cannot conclude anything from the univariate correlation until we control for all the relevant characteristics. Following that discussion, in the next section we present our identification strategy, which is based on the geographical inequalities associated with school resources across districts.

4) Methodological Approach

The previous section suggested the relevance of school resources on children's educational attainment, especially in contexts where these resources are scarce, as is the case of Peru. This section provides details on the methodological approach to allow us to separate the role of school resources *vis a vis* household and parent characteristics.

We estimate a multivariate model 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 test scores in Math (logic/mathematics) and Language (Integral Communication) for fourth graders from the 2001 ENRE. There are two key challenges in estimating this model. First, there could be unobserved district or school level characteristics that are correlated with both, school resources and socioeconomic status, and second, the selection problem, by which poorer students are also those sorting into low quality schools.

(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.

[21] It must be noted that Figure 4 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).

We first deal with the omitted variables bias without specific reference to the selection problem. If we understand that school environment is important, estimating an OLS model would 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 variables, then a random effects model at the school level will yield minimum variance estimates. Formally, the model to be estimated can be written as follows:

$$r_{ijk} = F_{ijk}\beta_1 + \sum_{n=1}^{3} Z_{jk}^{n}\gamma_n + D_k\beta_2 + \delta_{jk} + \varepsilon_{ijk} \qquad (1)$$

where r_{ijk} is a standardized score on the math or language test of student i, who attends to school j in district k. F_{ijk} is the vector of individual and household observable characteristics, D_k represent a vector of observable district characteristics, and $\sum_{n=1}^{3} Z_{jk}^{n}$ are three indicator variables that denoting school j's observable characteristics. Particularly, Z_{jk}^{n} indicates whether child i attend to a school of quality 1, 2, or 3. δ_{jk} 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 γ_n, which reflects the sign and statistical significance of the effect of the corresponding school level

---

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

[23] The necessary assumptions about the error term ε_{ijk}, and the random term δ_{jk} 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 (2007), Rivkin, Hanushek and Kain (2005). Clearly, a fixed effects model is more efficient at capturing the unobserved variation at the school level, while demanding less restrictive assumptions on the structure of the variance covariance matrix. However, we are interested in the relationship between school resources and school achievement, hence a fixed effect model at the school level will absorb all the relevant variation in the absence of more than one observation of school characteristics, which we do not have at this moment.

variables in explaining academic achievement. However, the shortcoming of the econometric estimation outlined in expression (1) is that we do not account for the fact that the household (parents) decide on the school to which child *i* will attend. This choice is shaped by parental preferences, but ultimately depends on the availability of suitable options. Richer or more concerned parents can often decide to send their children to a high or low quality schools. 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). Such selection would make students of different schools intrinsically different and would bias the effect of school resources.

In the ideal experiment, we would compare children randomly allocated to schools of different quality, and then compare their academic achievement. Given that such experiment is not possible (or available), we need to compare children that for reasons unrelated to their background, are forced to attend to schools of different quality. Public schools vary in their endowment of material and human resources, although political economy factors deviate the allocation of school resources at the primary school level from just following socio-economic differences. Hence, we can compare students with similar family and socioeconomic background, and while some were able to attend to a school with a quality chosen by their parents, others were restricted to attend the school type available in their district. Our identification strategy precisely accounts for these inequalities in the choice set to identify the effect of school resources on academic achievement, following a strategy similar to the one used in Contreras (2004). More precisely, we use the availability of classes of each type of school in district *k* as an instrument for the type of school to which child *i* attends.

In terms of interpretation, it is important to note that our two-stage procedure will estimate a local average treatment effect (LATE). This LATE will represent the causal effect of school quality on educational achievement for those children whose choices are affected by the availability of schools of different quality. The estimation of these causal effects are only going to be valid estimates for children who would have had a different choice of school had it been available. For example, we are not going to get any additional traction in our estimation from children who live in a district with only schools of type 1, but would have gone to that school *even if a better school was available*. Instead, if in that same district there were children who, in the presence of better schools, would have attended there, the estimator will capture exactly this variation.

The exclusion restriction for a valid instrument requires that the number of classrooms of each type of school in the district only affect academic achievement through the type of school to which the student attend. One concern that one might have is that the availability of schools of a certain quality is correlated with health and sanitary conditions in the district, which affects child health, and hence influence school achievement. To control for this possibility, we include in all regressions controls for the percentage of households with access to public sewage systems. Likewise, to control for the general wealth on the district, we introduce in the regression the proportion of households without access to public sewage and without public electricity. Once we include these variables, the conditional correlation between our instrument and the error term in the second stage regression due to district characteristics should disappear. We provide robustness checks for the exclusion restriction in the next section.

The idea is to instrument the effect of each type of school using information on the types of schools available on the districts where children from the ENRE 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. In 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.

As mentioned above, we form three groups of schools, depending on the school resources available and teacher's characteristics. We then count the number of classrooms of each type of school in each district and merge this district level dataset with the ENRE sample. This information is used in our two-stage methodology to compute, first, the effects of school availability on the type of school to which the child attends, and then the effect of this type of school on academic achievement. 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 (2):

$$Z^*_{ijk} = F_{ijk}\lambda_1 + NE1_k\lambda_2 + NE2_k\lambda_3 + NE3_k\lambda_4 + \delta_j + \mu_{ijk} \quad (2)$$

where Z^*_{ijk} is the school type of school *j*, where student *i*, resident of district *k*, attends. $NE1_k$, $NE2_k$ and $NE3_k$ represent the number of classrooms in schools of type 1, 2 and 3, operating in district *k*.[25] 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 plausible to assume that it is orthogonal to students' performance, their ability, or family unobservables (conditional on certain observables). However, the estimation is not done on Z^*, but on *Z*, thus we have:

$$\begin{aligned} &Z = 1 \text{ if } Z^* \leq 0 \\ &Z = 2 \text{ if } 0 \leq Z^* \leq \gamma \\ &Z = 3 \text{ if } \gamma \leq Z^* \end{aligned}$$

Assuming μ_{ijk} is normally distributed, in the first stage we estimate the probability that the school to which child *i* attends is of type 1, 2 or 3:

$$\begin{aligned} &\hat{Z}^1 = \Pr(Z = 1 \mid F, NE2, NE3) = \Phi(-F_{ijk}\hat{\gamma}_1 - NE2\hat{\gamma}_2 - NE3\hat{\gamma}_3) \\ &\hat{Z}^2 = \Pr(Z = 2 \mid F, NE2, NE3) = \Phi(\delta - F_{ijk}\hat{\gamma}_1 - NE2\hat{\gamma}_2 - NE3\hat{\gamma}_3) - \Phi(F_{ijk}\hat{\gamma}_1 - NE2\hat{\gamma}_2 - NE3\hat{\gamma}_3) \\ &\hat{Z}^3 = \Pr(Z = 3 \mid F, NE2, NE3) = 1 - \Phi(\delta - F_{ijk}\hat{\gamma}_1 - NE2\hat{\gamma}_2 - NE3\hat{\gamma}_3) \end{aligned}$$

Using these predicted probabilities, the second stage estimates the effect of school types on academic performance using a random effects model:

---

[25] 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.

$$r_{ijk} = F_{ijk}\beta_1 + \sum_{n=1}^{3} \hat{Z}_{jk}^{n}\gamma_n + D_k\beta_2 + \delta_{jk} + \varepsilon_{ijk} \qquad (3)$$

Comparing the estimates from equation (3) with those from equation (1) will allow us to estimate the relevance of supply side constraints on the academic achievement of Peruvian students. Notice that we control for observable district characteristics associated to socio-economic status, although we cannot discard that the omission of unobservable characteristics correlated with the availability of school types may bias our estimate. Nevertheless, we perform some robustness tests based on restricting the sample to those mostly affected by the availability of public schools, to show that the effect is indeed coming from the restrictions in school choice.

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 (1) through (4).

We first estimate the selection equation described in Equation (2). The dependent variable takes the value of 1, 2, or 3, depending on the type of school to which child *i* attend. Given the nature of the dependent variable, we use an ordered probit model, and include as independent variables the basic individual and family characteristics, and -more importantly- the availability classrooms of each type of school in the district. Table 3 reports the relevant coefficients for the sample of students enrolled in fourth grade who took the Math and Language tests.[26] Family characteristics, such as the years of schooling of the most educated member of the household, the student's 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

[26] There are more students who took the Math than Language test. Table A. 4 shows the marginal coefficients associated with the regressions shown in Table 3.

include in the regression the percentage of households in the district without access to sewage and without electricity. These variables proxy for the wealth level of the district, and makes sure that the estimators obtained for the supply of classrooms in the district are not capturing other observable characteristics in the locality. We find that living in a district with more population without electricity decreases the probability of attending to a high quality school, while more people without sewage does not affect that probability at all.

Importantly, the results in Table 3 confirm that the availability of quality schools in the district, as measured by the number of classrooms in each type of school available in the district significantly affects the school the child attends. There is a negative coefficient for the number of classrooms of schools of type 1 and type 2 in the district meaning it reduces the probability of the child attending a school of type 3. On the other hand, the presence of classrooms of schools type 3 increase the probability that a child attends a high quality school. These results are statistically and economically significant for the sample of children who took the Math and Language test. 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.

The second stage consists in the estimation of equation (3), where we show the causal effect of school resources (instrumented by school availability) on academic achievement. Table 4 shows the results for the standardized test scores in Math and Language, and compares them to the ones obtained in using a simple random effects model (RE, as outlined in equation (1)) and the two stage random effects model described above (RE-IV, as described in equation (4)). 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. If the effect estimated when we control for the school choice is lower than the one assuming random distribution of school resources (RE), children from more educated or richer parents are the ones who are sent to high quality schools, and hence the ones who have higher educational attainment. In this case, 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 equitable distribution of school resources.[27]

The results shown in Table 4 are consistent with the second hypothesis. When we estimate the RE model disregarding the endogeneity of the school choice, the estimated effects of school characteristics are relatively small, and only significant for the high quality schools in the Math and Language tests (low quality schools are the omitted category). Attending to a high quality increases school attainment in 0.28 in the Math and Language tests.[28] The characteristics of the intermediate type of schools do not appear to affect student's performance, when compared to the poorer type of schools.

In Columns (2) and (4), we use our two-stage method to correct for selection bias associated to the geographical distribution of quality schools. Controlling for the limited supply of schools, attending to a medium or high quality school has a positive and significant effect on students' performance. These effects are an order of magnitude larger than the one estimated in Columns (1) and (3), when we ignored the selection bias. The estimated effect of attending to a medium quality school goes from almost zero to 0.35SD in Math, while for Language, the effect does not seem to be statistically relevant. The effect for high quality schools increases from 0.28SD in both cases to 1SD in Math and 0.52SD in Language.

As for the child and family characteristics included in our models, boys are more likely to perform better in Math. Native speakers have lower educational attainment, while those from more educated and wealthier parents tend to perform better at school. However, when we take into account the endogeneity school choice, the estimated coefficient for parents' education and the household asset index drop significantly, both in magnitude and statistical significance, and this is particularly the case for Math scores. These patterns

---

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

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

reinforce the intuition behind our instrumental variable approach. If we fail to control for the availability of schools, we would be wrongly attributing the effect of school resources to household and parent characteristics.

School resources might be more beneficial for particular groups of the population. Given the current trends of gender inequalities in education, which show that girls are attending more often to school than boys, and performing better, it is particularly important to see if these changes can be partially accounted for by school resources. Table 6 presents the coefficients of interest of the RE-IV regressions, disaggregated by gender.[29] High quality schools seem to benefit girls more than boys, both in Language and Math. After taking into account the constraints in school supply, it appears that the effect of school characteristics on academic performance is mostly driven by girls, and somewhat by its effect on boys, but just in Math. A 1SD increase in the probability of attending to a high quality school improves girls' standardized test scores in both Mathematics and Language in about 1.3SD and 1.05SD, respectively, whereas for the boys an increase in this probability does increases their Math test score by 0.6SD, and does not appear to improve their Language test score. Therefore, an improvement in the quality of local school supply significantly contributes to closing the gender gap in educational performance.

In Table 5 we check the robustness of our results, and show the RE and RE-IV results of the main regressions shown in Table 4 for different subgroups of the population. Particularly, we focus our attention on the richest and the poorest, as measured by the wealth in the district or in the household. The results show that the increase in the magnitude of the coefficients associated with the type of school to which the child attends only takes place among the most constrained students. For example, among those who live in districts in the 2 poorest quintiles (see Figure **3**) the relevance of school characteristics is an order of magnitude larger once we account for the limited availability of schools, and the result holds for schools of type 2 and 3, in Math and Language. Conversely, when we look at children living in richer districts (in which there are only schools of type 2 and 3), the

[29] Table A. 5 shows the first stage of these regressions. The results are very similar to those presented for the overall student population in Table 3.

quality of the school to which the child attends is zero, both with and without instrumenting. This result is consistent with the non-linearity hypothesis suggested in the literature.[30] These patterns are similar when we compare the results for the poorest and richest, as measured by the asset index. These results are not as sharp as those shown in the previous panels, since we can have poor households living in unconstrained districts, as well as rich households living in constrained districts.

The evidence shown in Tables 3 through 5 provides empirical support for the hypothesis that geographic inequalities in the distribution of elementary schools with adequate resources is a serious barrier to overcoming inequalities in education in Peru, and this result is especially relevant for girls. Moreover, this barrier is 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.

The coefficients estimated in Table 4 and 5 allow us to asses the average effect of each factor to academic achievement, but are not helpful if we want to analyze how do they affect students in different parts of the distribution, since their impact will not only depend on the coefficients, but also on the other variables that contribute to differences between sub-groups and measurement scales. Particularly, we are interested in measuring the extent to which the differences in educational attainment between children from rich and poor districts can be explained by the variation in school characteristics. Manipulating equation (1) from the previous section, we are able to decompose the differences in educational attainment between children from the richest quintile (V) and the poorest one (I) as follows:[31]

$$1=\sum_{j}\frac{\beta_{1j}(\overline{F}_{Vj}-\overline{F}_{Ij})}{(\overline{r}_{V}-\overline{r}_{I})}+\sum_{k}\frac{\gamma_{1k}(\overline{Z}_{Vj}-\overline{Z}_{Ij})}{(\overline{r}_{V}-\overline{r}_{I})}+\sum_{k}\frac{(\overline{\delta}_{Vj}-\overline{\delta}_{Ij})}{(\overline{r}_{V}-\overline{r}_{I})}+\sum_{k}\frac{\beta_{2k}(\overline{D}_{Vk}-\overline{D}_{Ik})}{(\overline{r}_{V}-\overline{r}_{I})}+\frac{(\overline{\varepsilon}_{Vk}-\overline{\varepsilon}_{Ik})}{(\overline{r}_{V}-\overline{r}_{I})} \quad (4)$$

where the variables represent the same as in equation (1) and the bars denote averages by quintiles (I and V, denoted in the subindices). Each term on equation (4) represents the

[30] See: Hanushek (2003), Fuller and Heyneman (1989), Baker and LeTendre (2002), Baker et al (2001), Chudgar and Luschei (2009 and 2011).

[31] See Valdivia (2002).

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 the coefficients, 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. Solving (4) for each variable included in the analysis allows us to assess the exact contribution of each variable to the differences between extreme quintiles. Table 7 shows the results from this decomposition, using the results from our simple random effects model and from the one in which we account for the endogeneity of school choice (RE-IV).

Columns (1) and (3) in Table 7 show the contribution to each observable (and unobservable) factor to the differences in academic achievement between children in poor and rich districts. While child and family characteristics account for 35 and 48 percent of the differences in Math and Language, respectively. School characteristics only contribute to 15 and 14 percent of those differences, while district characteristics are responsible for 34 and 30 percent of them. When implementing our correction for supply constraints, as shown in Columns (2) and (4), the importance of household characteristics to these differences is substantially reduced. School resources now account for 32 and 23 percent of them in Math and Language, respectively. In both fields, the highest contribution to inequality comes from high quality schools. This increase comes at the expense of the contribution of home inputs, especially of the SES indicator and parental education. The portion of the inequality explained by unobserved factors marginally increases also, particularly in Math. In other words, failing to control for the constraints that geographic inequalities in the distribution of school resources 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.

Table 8 shows the results from a similar exercise based on the results from Table 7, where we disaggregated the analysis by gender. Once we control for the conditional probability of attending to each type of school, the contribution of the school variables more than doubles, and again this happens at the expense of the household characteristics.

This result in much more pronounced for girls than for boys, suggesting that improving the quality of the schools at the local level not only will help to close the gender gap in schooling, but also will significantly reduce the socioeconomic gaps in attainment.

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. Using information from public schools in Peru, our 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 percent. 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 15 percent to 32 percent (14 percent to 23 percent in language). These results are more pronounced for girls than for boys, suggesting that an improvement of school characteristics will naturally have a gender bias, favoring girls.

The literature review provided in the paper illustrates the long discussion on the ways that school characteristics determine the educational quality. Most of the literature for developed countries concludes 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 is a significant effect of school resources on academic achievement. The evidence from Peruvian public schools supports this later interpretation. 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. However, if children from more educated or richer parents, or from parents who are more concerned about the quality of the education, 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 in their localities, 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 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, they are heterogeneous in terms of the physical and human resources available, such as qualified teachers, school materials, and equipment. Previous estimates of the school effects 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

Andrabi, Tahir, Jishnu Das and Asim Khwaja (2008). “Dime a Day: The Possibilities and Limits of Private Schooling in Pakistan.” Comparative Education Review, vol. 52(3): 329-355

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, Vol. 114(2): 533-575.

Baker, D., Goesling, B. and LeTendre, G. (2002). “Socio-economic Status, School Quality, and National Economic Development: A Cross-national Analysis of the “Heyneman-Loxley Effect” on Mathematics and Science Achievement” Comparative Education Review, Vol. 46 (3): 291-312.

Baker, D. , Akiba, M., LeTendre, G., and Wiseman, A. (2001). “Worldwide Shadow Education: Outside-School Learning, Institutional Quality of Schooling, and Cross-national Mathematics Achievement.” Education, Evaluation, and Policy Analysis, Vol. 23(1): 1-17.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo and Leigh Linden (2007). “Remedying Education: Evidence from Two Randomized Experiments in India.” The Quarterly Journal of Economics (2007) 122 (3): 1235-1264

Barnett, Steven (2003). “Better Teachers, Better Preschools: Student Achievement Linked to Teacher Qualifications.” NIEER Preschool PolicyMatters, Issue 2.

Behrman, Jere R., Susan W. Parker, and Petra E. Todd (2011). "Incentives for Students and Parents." Paper presented at Education Policy in Developing Countries: What Do We Know, and What Should We Do to Understand What We Don’t Know?, February 4□ 5, at University of Minnesota.

Buddin, Richard and Gema Zamarro (2009). “Teacher qualifications and student achievement in urban elementary schools.” *Journal of Urban Economics*, Vol. 66(2): 103–115

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

Card, David; Alan B. 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, and Edward Vytlacil (2003). “Understanding what instrumental variables estimate: Estimating marginal and average returns to education”. Manuscript, University of Chicago, July.

Case, Anne, and 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.

Chudgar, Amita and Thomas F. Luschei (2009). “National Income, Income Inequality, and the Importance of Schools: A Hierarchical Cross-National Comparison” American Educational Resesearch Journal, vol. 46(3): 626-658

…………. (2011). “Teachers, Student Achievement and National Income: A cross national examination of relationships and interactions” Prospects, Vol. 41: 507-533

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor (2010). “Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects” Journal of Human Resources, Vol. 45(3): 655-681

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert York (1966). "Equality of educational opportunity." Washington, DC.

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

Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman (2013). “School Inputs, Household Substitution, and Test Scores.” American Economic Journal: Applied Economics, Vol. 5(2): 29–57

Duflo, Esther (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." American Economic Review, 91(4): 795-813.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2014). “School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools” Stanford University WP.

Fuller and Heyneman (1989) “Third World School Quality: Current Collapse, Future Potential.” Educational Researcher, Vol. 18: 12-19.

Galiani, S; P. Gertler, and E. Schargrodsky (2008). “Helping the Good Get Better, but Leaving the Rest Behind: How Decentralization Affects School Performance.” Journal of Public Economics 92 (10-11): 2106-2120 .

Gamoran, Adam and Daniel A. Long (2006). “Equality of Educational Opportunity: A 40-Year Retrospective” WCER Working Paper No. 2006-9

Gamoran, Adam (2001). “American Schooling and Educational Inequality: A Forecast for the 21st Century.” Sociology of Education, Vol. 74, Extra Issue: Current of Thought: Sociology of Education at the Dawn of the 21st Century: 135-153

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, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz (2004). "Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya." Journal of Development Economics Vol. 74(1): 251-268.

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.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin (2009). "Many Children Left Behind? Textbooks and Test Scores in Kenya." American Economic Journal: Applied Economics, Vol. 1(1): 112-35.

Glewwe, Paul, Eric A. Hanushek, Sarah Humpage, and Renato Ravina (2011). “School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010” Paper presented at Education Policy in Developing Countries: What Do We Know, and What Should We Do to Understand What We Don’t Know?, February 4-5, at University of Minnesota.

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

Guerrero, Gabriela (2010). "Del dicho al hecho hay mucho trecho: una análisis de la implementación de la Política Nacional de Educación Bilingüe Intercultural en Puno, Peru". In: Benavides, Martín; Neira, Paul (eds). Cambio y continuidad en la escuela peruana: una mirada institucional a la implementación de programas, procesos y proyectos educativos. Lima: GRADE. p. 55-118.

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”. World Bank Research Observer 10 (2): 227–246.

Heyneman, Stephen P., and William A. Loxley. 1982. "Influences on Academic Achievement across High and Low Income Countries: A Re-analysis of IEA Data." Sociology of Education, Vol. 55 (1): 13-21.

…………... (1983a). "The Distribution of Primary School Quality within High and Low Income Countries." Comparative Education Review.

………….. (1983b). “The Effect of Primary-School Quality on Academic Achievement Across Twenty-nine High-and Low-Income Countries.” American Journal of Sociology, Vol. 88(6): 1162-1194

Jamison, Dean T., Barbara Searle, Klaus Galda, and Stephen P. Heyneman (1981). "Improving elementary mathematics education in Nicaragua: An experimental study of the impact of textbooks and radio on achievement." Journal of Educational Psychology Vol. 73(4): 556.

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.

Kremer, Michael and Edward Miguel (2004). “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” Econometrica, Vol. 72(1): 159–217

Kremer, Michael, Conner Brannen, and Rachel Glennerster (2013). “The Challenge of Education and Learning in the Developing World.” Science, Vol. 340(6130): 297-300

McEwan, P., M. Urquiola, and E. Vegas (2008) “School Choice, Stratification, and Information on School Perfomance: Lessons From Chile.” Economía (Journal of the Latin American Economic Association), Vol. 8(2): 28-42.

Pal, Sarmistha (2009). “Public Infrastructure, Location of Private Schools and Primary School Attainment in an Emerging Economy.” IZA WP 4572

Pritchett, Lant, and Deon Filmer (1999). “What Education Production Functions Really Show: A Positive Theory of Education Expenditures.” Economics of Education Review, Vol. 18(2): 223–39.

……….. (2001). "Estimating Wealth Effects without Expenditure of Data - Or Tears: An Application to Enrollments in States of India". Demography.

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

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

Tan, Jee-Peng, Julia Lane, and Paul Coustere (1997). "Putting inputs to work in elementary schools: What can be done in the Philippines?." Economic development and cultural change Vol. 45(4): 857-879.

Todd, Petra E.; Kenneth I. Wolpin (2007). “The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps”. *Journal of Human Capital* 1: 91-136.

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

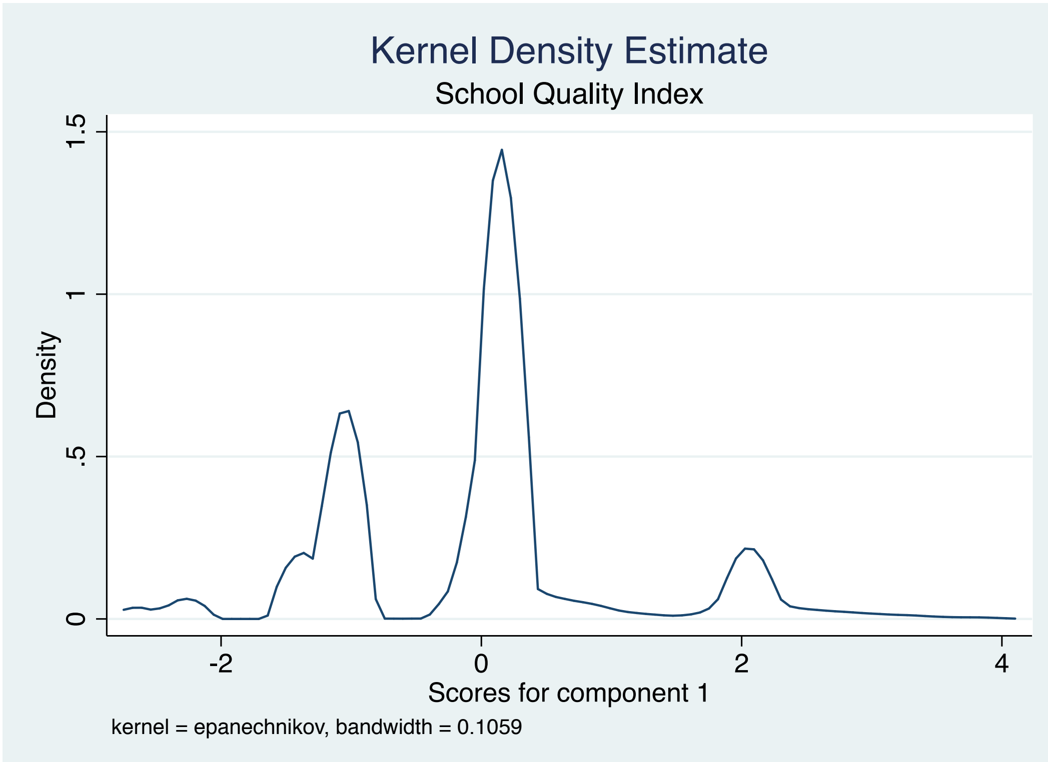
Torreblanca, A.; D. Zacharías (2002b). “*Diseño muestral Evaluación Nacional 2001: Ficha técnica-secundaria*”. In: http://www.minedu.gob.pe/mediciondelacalidad/2003/

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

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.

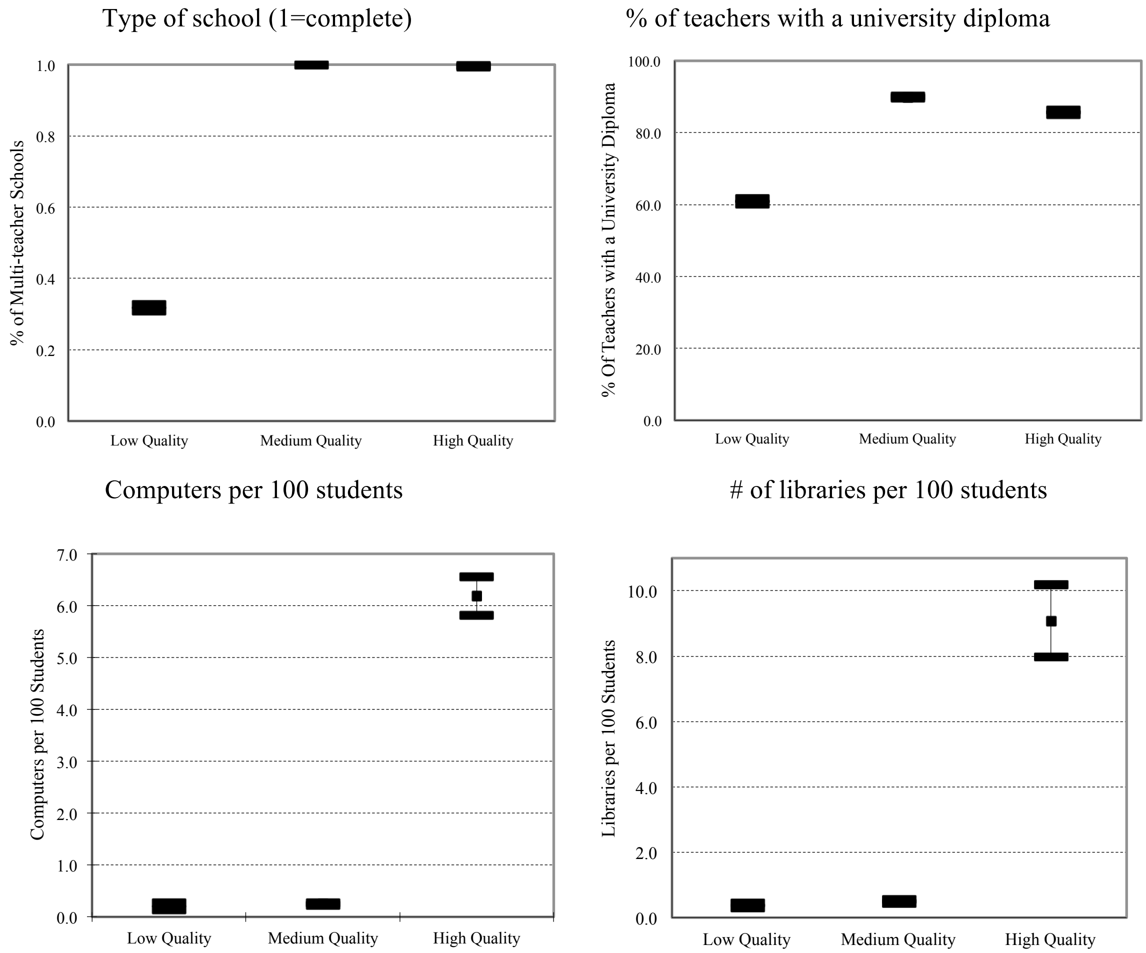
Yamada, Gustavo (2010). “Migración interna en el Peru”. Documento de Trabajo # 86, Universidad del Pacífico, Lima.

Figure 1: Empirical Distribution of the School Quality Index



Note: The figure shows the kernel density estimate for the School Quality index generated using the universe of public schools in Peru.

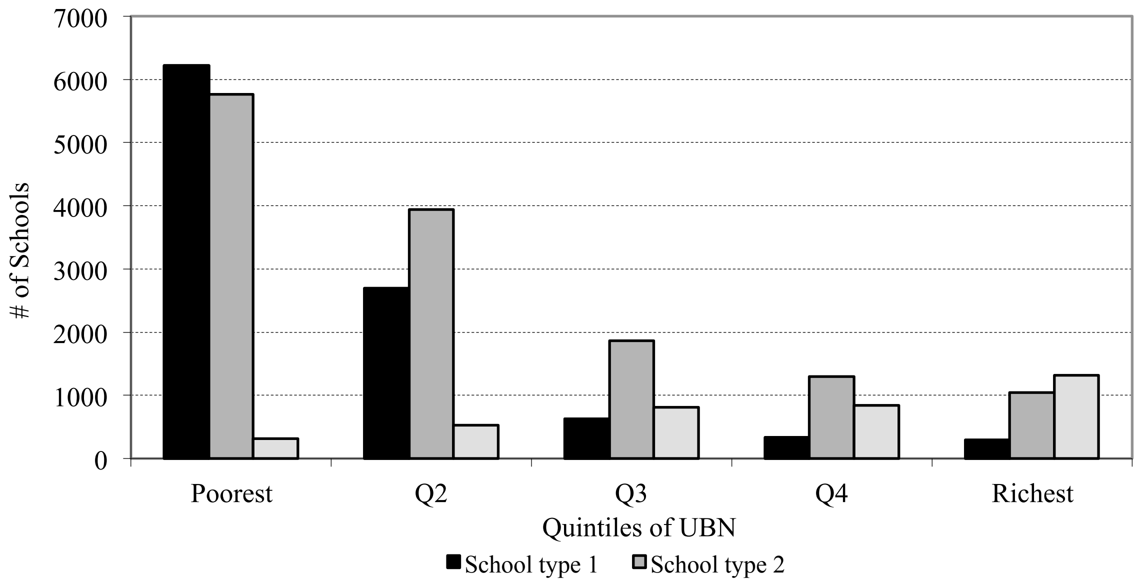
Figure 2: Averages and Confidence intervals of variables associated with School Type.



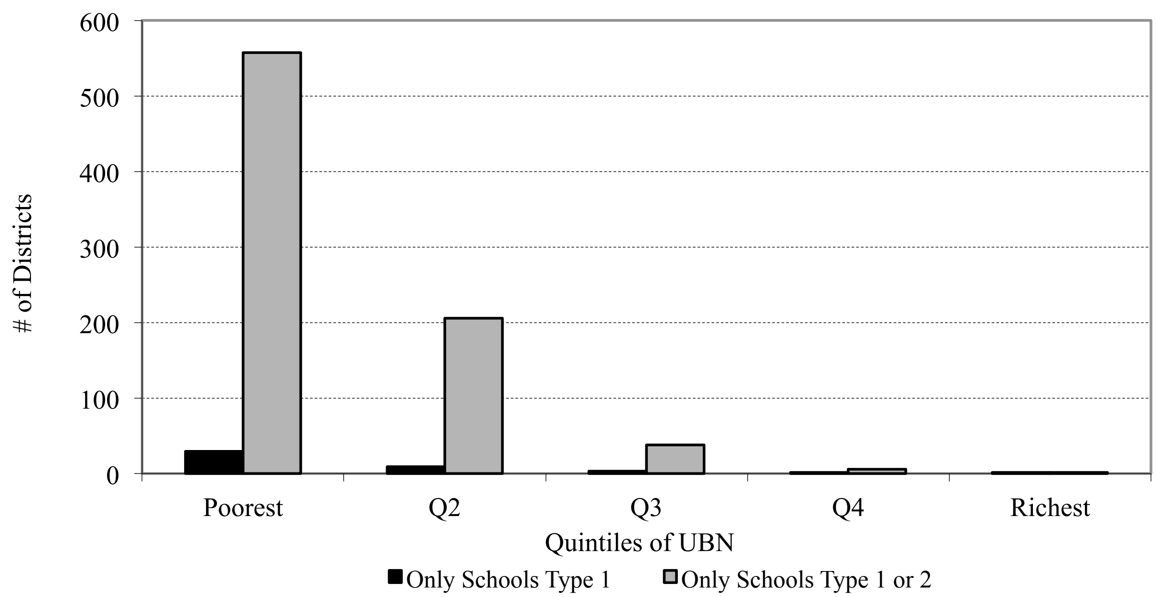
Source: National School Census.
Notes: The Figure shows the mean and 95% confidence interval for each of the variables considered in the school quality index, for each type of school. The sample includes the universe of public schools in Peru. 3,807 schools are classified as high quality, 13,935 medium quality, and 10,180 are low quality.

Figure 3: Number of districts with choice constraints, by Quintiles of UBN (population weighted)

Panel A: Availability of schools of each type, by district wealth

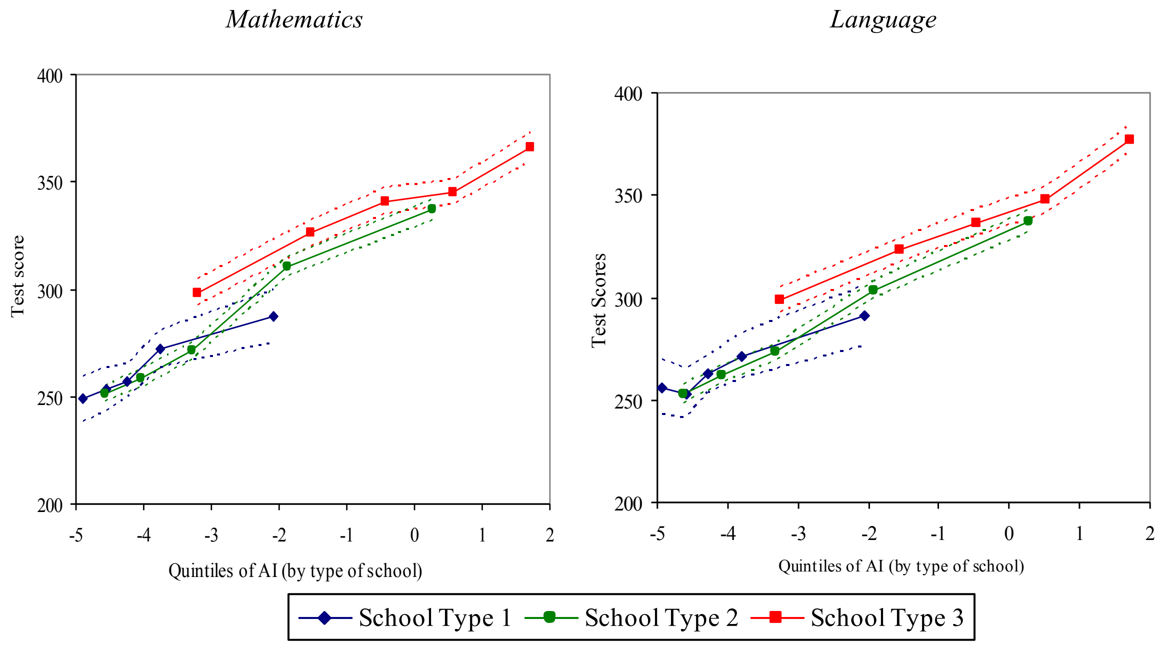


Panel B: Constraints in supply of schools of each type, by district wealth



Source: National School Census.
Notes: Panel A shows the number of public schools of each type available in districts, by quintiles of unsatisfied basic needs. Panel B shows the number of districts in which there are only some types of schools.

Figure 4: Differences on school attainment in elementary, by school type and SES



Source: Authors elaboration using information from ENRE 2001.

Table 1: Descriptive statistics, variables used in the regression analysis

	Description	Source
Logics and mathematics		
Standardized score	Standardized test score in Math and Language. The original scores are normalized	Standardized test. ENRE 2001
Child's Gender	Student's gender, =1 if boy.	Student's questionnaire. ENRE 2001
Child's mother tongue	Language that the child commonly uses at home. =1 if Quechua, Aymara, or other native language.	
Highest level of schooling in the HH	Highest level of schooling between the care takers (eg. Mother and father). 1=No education; 2=Incomplete Primary; 3=Complete Primary; 4=Incomplete Secondary; 5=Complete Secondary; 6=More than secondary	Household questionnaire. ENRE 2001
HH asset index	Asset index computed using information on the dwelling's characteristics and asset tenancy. See Table A. 3 for descriptive statistics on the variables considered, and Section3 for a description of the construction of the AI.	
School type	School classification. =1 if each teacher takes care of only one class.	2002 School Census.
% of teachers with a university diploma in the school	% teachers in the school who have completed a university degree	
Operative computers per 1000 students	Number of operative computers in the school, divided by the number of students enrolled (*1000 for scaling)	
# of libraries in the school per 1000 students	Number of operative libraries in the school, divided by the number of students enrolled (*1000 for scaling purposes)	
% of pop. w/o sewage in the district	% of households in the district that do not have a connection to public sewage system.	1993 National Population Census
% of pop. w/o electricity in the district	% of households in the district that do not have a connection to public electricity system.	

Table 2: Distribution of the number of public schools by quintiles of UBN (%) - ENRE 2001 vs. School Census 2002

Quintiles of UBN	ENRE2001 Total	ENRE2001 (1)	ENRE2001 (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 3: Ordered probit for the decision of child's school – 4th grade

	Type of school	
	Mathematics	Language
Child's Gender (1= boy)	-0.040	-0.026
	(0.044)	(0.046)
Child's mother tongue (1= native)	-0.673	-0.705
	(0.099)***	(0.100)***
Highest level of schooling in the HH	0.071	0.064
	(0.028)**	(0.028)**
HH asset index	0.202	0.205
	(0.035)***	(0.035)***
# of classrooms in schools type 1 in the district (divided by 100)	-0.989	-0.984
	(0.256)***	(0.255)***
# of classrooms in schools type 2 in the district (divided by 100)	-0.136	-0.144
	(0.064)**	(0.064)**
# of classrooms in schools type 3 in the district (divided by 100)	0.145	0.145
	(0.038)***	(0.037)***
% of pop. w/o sewage in the district	-0.131	-0.116
	(0.259)	(0.260)
% of pop. w/o electricity in the district	-0.420	-0.431
	(0.229)*	(0.228)*
Observations	4,675	4,098
Pseudo R-sq	524	519
Chi^2	0.28	0.28
Log Likelihood	283.59	285.42

Coefficients from ordered probit regressions reported, marginal coefficients for each category are shown in the appendix. Robust standard errors clustered at the school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Source: ENRE 2001, School census 2000.

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

	Mathematics		Language	
	RE	RE-IV	RE	RE-IV
Child's Gender (1= boy)	0.145	0.155	-0.031	-0.028
	(0.020)***	(0.021)***	(0.022)	(0.022)
Child's mother tongue (1= native)	-0.140	0.002	-0.243	-0.188
	(0.032)***	(0.058)	(0.037)***	(0.060)***
Highest level of schooling in the HH	0.044	0.028	0.048	0.042
	(0.009)***	(0.010)***	(0.009)***	(0.010)***
HH asset index	0.083	0.036	0.120	0.095
	(0.011)***	(0.018)*	(0.013)***	(0.020)***
Attends to school type 2	0.069		0.064	
	(0.081)		(0.071)	
Attends to school type 3	0.282		0.289	
	(0.098)***		(0.084)***	
Predicted probability attends to school type 2		0.357		0.055
		(0.194)*		(0.186)
Predicted probability attends to school type 3		1.060		0.519
		(0.333)***		(0.292)*
% of pop. w/o sewage in the district	-0.571	-0.510	-0.535	-0.521
	(0.127)***	(0.136)***	(0.106)***	(0.114)***
% of pop. w/o electricity in the district	-0.192	-0.115	-0.179	-0.183
	(0.109)*	(0.121)	(0.089)**	(0.104)*
Constant	0.259	-0.331	0.382	0.246
	(0.124)**	(0.301)	(0.109)***	(0.252)
Observations	4,675	4,675	4,098	4,098
Number of schools	524	524	519	519
R sq.	0.34	0.33	0.42	0.42
ρ	0.39	0.39	0.28	0.28
Chi^2	577.19	571.08	964.73	954.97

Robust standard errors clustered at the school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Source: ENRE 2001, School census 2002.

Table 5: Robustness Check – Determinants of Academic Achievement in Different Samples

	Mathematics		Language	
	RE	RE-IV	RE	RE-IV
Full Sample				
Attends to school type 2	0.069		0.064	
	(0.081)		(0.071)	
Attends to school type 3	0.282		0.289	
	(0.098)***		(0.084)***	
Predicted probability attends to school type 2		0.357		0.055
		(0.194)*		(0.186)
Predicted probability attends to school type 3		1.06		0.519
		(0.333)***		(0.292)*
2 Poorest UBN Quintiles				
Attends to school type 2	0.025		0.005	
	(0.082)		(0.069)	
Attends to school type 3	0.202		0.282	
	(0.115)*		(0.096)***	
Predicted probability attends to school type 2		0.321		-0.098
		(0.211)		(0.153)
Predicted probability attends to school type 3		0.859		0.342
		(0.391)**		(0.319)
2 Richesy UBN Quintiles				
Attends to school type 2	-		-	
Attends to school type 3	0.010		-0.018	
	(0.091)		(0.080)	
Predicted probability attends to school type 2		-		-
Predicted probability attends to school type 3		-0.007		-0.077
		(0.215)		(0.230)
2 Poorest AI Quintiles				
Attends to school type 2	0.010		-0.001	
	(0.083)		(0.073)	
Attends to school type 3	0.264		0.424	
	(0.136)**		(0.121)***	
Predicted probability attends to school type 2		0.179		-0.217
		(0.240)		(0.167)
Predicted probability attends to school type 3		1.369		1.449
		(0.358)***		(0.362)***
2 Richest AI Quintiles				
Attends to school type 2	0.368		0.304	
	(0.339)		(0.346)	
Attends to school type 3	0.443		0.366	
	(0.341)		(0.348)	
Predicted probability attends to school type 2		0.913		1.838
		(4.452)		(4.482)
Predicted probability attends to school type 3		1.069		1.572
		(4.341)		(4.339)

Selected coefficients from regressions similar to the ones shown in Table 3. Robust standard errors clustered at the school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Determinants of school attainment (4th grade) , by gender (selected coefficients)

	Mathematics	Language
Girls		
Predicted probability attends to school type 2	0.375	0.243
	(0.310)	(0.235)
Predicted probability attends to school type 3	1.336	1.053
	(0.429)***	(0.360)***
Observations	2,321	2,043
Number of schools	501	491
R sq.	0.35	0.45
Boys		
Predicted probability attends to school type 2	0.067	-0.129
	(0.244)	(0.218)
Predicted probability attends to school type 3	0.593	-0.009
	(0.330)*	(0.310)
Observations	2,354	2,055
Number of schools	496	485
R sq.	0.33	0.40

Robust standard errors clustered at the school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Source: ENRE 2001, School census 2002.
Regressions are similar to those shown in Columns 2 and 4 of Table 3. Includes controls for Child's mother tongue, highest schooling in the household, HH asset index, % of pop. w/o sewage, and w/o electricity. The full table is available upon request.

Table 7: Contribution of each variable to the inequalities in academic achievement between UBN quintiles

	Mathematics		Language	
	RE	RE-IV	RE	RE-IV
Child and household characteristics	*34.99*	*14.37*	*47.44*	*38.05*
Child's Gender (1= boy)	-0.60	-0.64	0.12	0.11
Child's mother tongue (1= native)	4.12	-0.06	6.62	5.12
Highest level of schooling in the HH	7.02	4.47	7.22	6.31
HH asset index	24.44	10.60	33.48	26.50
School	*15.44*	*32.27*	*14.15*	*23.21*
Attends to school type 2	-1.80		-1.60	
Attends to school type 3	12.60		12.24	
Predicted probability attends to school type 2		-10.32		-1.52
Predicted probability attends to school type 3		48.51		22.47
μ_i (School-level unobservables)	4.64	-5.93	3.52	2.26
District	*33.93*	*27.88*	*30.12*	*29.68*
% of pop. w/o sewage in the district	25.68	22.94	22.83	22.23
% of pop. w/o electricity in the district	8.25	4.94	7.29	7.46
Unexplained	*15.64*	*25.48*	*8.29*	*9.05*
Total	*100.00*	*100.00*	*100.00*	*100.00*

Source: Calculations based on results from Table 3.

Table 8: Contribution of each group of variables to the inequalities in academic achievement between welfare UBN and gender

	Mathematics		Language	
	RE	RE-IV	RE	RE-IV
Girls				
Child and household characteristics	*39.59*	*11.58*	*54.65*	*38.51*
School	*20.06*	*49.39*	*13.17*	*34.71*
District	*30.24*	*22.85*	*27.98*	*24.37*
Unexplained	*10.11*	*16.19*	*4.20*	*2.41*
Total	*100.00*	*100.00*	*100.00*	*100.00*
Boys				
Child and household characteristics	*44.32*	*34.10*	*52.39*	*55.94*
School	*15.93*	*24.55*	*-2.86*	*4.67*
District	*28.23*	*27.76*	*27.34*	*30.90*
Unexplained	*11.53*	*13.60*	*23.12*	*8.49*
Total	*100.00*	*100.00*	*100.00*	*100.00*

Source: Calculations based on results from Table 4.

APPENDIX (Not Intended for Publication)

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

	Obs.	Mean	S. D.	Min	Max
Logics and mathematics					
Standardized score	4675	-0.07	0.99	-3.53	3.93
Child's Gender (1= boy)	4675	0.50	0.50	0.00	1.00
Child's mother tongue (1= native)	4675	0.26	0.44	0.00	1.00
Highest level of schooling in the HH	4675	3.97	1.61	1.00	6.00
HH asset index	4675	-2.40	2.02	-5.38	2.85
School type (1=multi-teacher)	4675	0.43	0.50	0.00	1.00
% of teachers with a university diploma in the school	4675	83.15	21.97	0.00	100.00
Operative computers per 100 students	4675	0.81	2.81	0.00	31.75
# of libraries in the school per 100 students	4675	0.67	2.21	0.00	27.78
Attends to school type 1	4675	0.14	0.35	0.00	1.00
Attends to school type 2	4675	0.51	0.50	0.00	1.00
Attends to school type 3	4675	0.34	0.48	0.00	1.00
# of classrooms in schools type 1 in the district (divided by 100)	4675	0.25	0.33	0.00	1.64
# of classrooms in schools type 2 in the district (divided by 100)	4675	1.49	1.99	0.00	12.39
# of classrooms in schools type 3 in the district (divided by 100)	4675	1.96	3.79	0.00	20.35
% of pop. w/o sewage in the district	4675	0.64	0.30	0.00	1.00
% of pop. w/o electricity in the district	4675	0.49	0.34	0.00	1.00
Integral Communication					
Standardized score	4098	-0.08	1.00	-2.92	3.29
Child's Gender (1= boy)	4098	0.50	0.50	0.00	1.00
Child's mother tongue (1= native)	4098	0.26	0.44	0.00	1.00
Highest level of schooling in the HH	4098	3.99	1.61	1.00	6.00
HH asset index	4098	-2.34	2.02	-5.34	2.73
School type (1=multi-teacher)	4098	0.42	0.49	0.00	1.00
% of teachers with a university diploma in the school	4098	83.37	21.76	0.00	100.00
Operative computers per 100 students	4098	0.80	2.77	0.00	31.75
# of libraries in the school per 100 students	4098	0.66	2.15	0.00	27.78
Attends to school type 1	4098	0.14	0.35	0.00	1.00
Attends to school type 2	4098	0.51	0.50	0.00	1.00
Attends to school type 3	4098	0.35	0.48	0.00	1.00
# of classrooms in schools type 1 in the district (divided by 100)	4098	0.25	0.33	0.00	1.64
# of classrooms in schools type 2 in the district (divided by 100)	4098	1.52	2.03	0.00	12.39
# of classrooms in schools type 3 in the district (divided by 100)	4098	2.00	3.85	0.00	20.35
% of pop. w/o sewage in the district	4098	0.64	0.30	0.00	1.00
% of pop. w/o electricity in the district	4098	0.48	0.34	0.00	1.00

Sources: ENRE 2001, National School Census 2002, and Population Census 1993. The index that classifies schools into 3 types is constructed using a principal components analysis, which includes (i) % of teachers with a university diploma; (ii) an indicator for whether the school is complete or one-teacher/multigrade school; (iii) # of computers per 1000 students and (iv) # of libraries available per 1000 students.

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

	Coefficient	Mean	Std. Dev	Min	Max
School type (1=multi-teacher)	0.719	0.8	0.4	0.0	1.0
% of teachers with a university diploma in the school	0.307	78.8	32.3	0.0	100.0
Operative computers per 100 students	0.601	1.0	5.3	0.0	221.0
# of libraries in the school per 100 students	0.165	1.6	13.7	0.0	1000.0
% of the overall variance explained by the first PC	28.24%				

Table A. 3: Asset index principal component analysis

Variable		Math Sample	Language 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				
	Dependent variable: Type of school			
	Mathematics		Language	
	Coef.	Std. Err.	Coef.	Std. Err.
Pr(E=1 | X)				
Child's Gender (1= boy)	0.005	0.005	0.003	0.005
Child's mother tongue (1= native)	0.101	0.022	0.103	0.022
Highest level of schooling in the HH	-0.008	0.003	-0.007	0.003
HH asset index	-0.023	0.005	-0.023	0.005
# of classrooms in schools type 1 in the district (divided by 100)	0.114	0.032	0.109	0.031
# of classrooms in schools type 2 in the district (divided by 100)	0.016	0.007	0.016	0.007
# of classrooms in schools type 3 in the district (divided by 100)	-0.017	0.004	-0.016	0.004
% of pop. w/o sewage in the district	0.015	0.030	0.013	0.029
% of pop. w/o electricity in the district	0.048	0.026	0.048	0.025
Pr(E=2 | X)				
Child's Gender (1= boy)	0.009	0.010	0.006	0.010
Child's mother tongue (1= native)	0.100	0.017	0.107	0.018
Highest level of schooling in the HH	-0.016	0.006	-0.014	0.007
HH asset index	-0.044	0.009	-0.047	0.009
# of classrooms in schools type 1 in the district (divided by 100)	0.217	0.057	0.223	0.059
# of classrooms in schools type 2 in the district (divided by 100)	0.030	0.015	0.033	0.015
# of classrooms in schools type 3 in the district (divided by 100)	-0.032	0.009	-0.033	0.009
% of pop. w/o sewage in the district	0.029	0.057	0.026	0.059
% of pop. w/o electricity in the district	0.092	0.052	0.098	0.054
Pr(E=3 | X)				
Child's Gender (1= boy)	-0.013	0.015	-0.009	0.016
Child's mother tongue (1= native)	-0.200	0.027	-0.210	0.027
Highest level of schooling in the HH	0.024	0.009	0.021	0.010
HH asset index	0.068	0.012	0.069	0.012
# of classrooms in schools type 1 in the district (divided by 100)	-0.331	0.083	-0.332	0.084
# of classrooms in schools type 2 in the district (divided by 100)	-0.045	0.022	-0.049	0.022
# of classrooms in schools type 3 in the district (divided by 100)	0.049	0.013	0.049	0.013
% of pop. w/o sewage in the district	-0.044	0.086	-0.039	0.088
% of pop. w/o electricity in the district	-0.141	0.077	-0.145	0.077

Robust standard errors, clustered at the school level, in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Source: Calculations based on results from Table 3.

Table A. 5: Ordered probit for the decision of child's school – 4th grade

	Type of school			
	Girls		Boys	
	Mathematics	Language	Mathematics	Language
Child's mother tongue (1= native)	-0.720	-0.740	-0.628	-0.673
	(0.117)***	(0.118)***	(0.108)***	(0.108)***
Highest level of schooling in the HH	0.089	0.084	0.052	0.041
	(0.034)***	(0.035)**	(0.029)*	(0.030)
HH asset index	0.188	0.187	0.220	0.227
	(0.040)***	(0.040)***	(0.040)***	(0.041)***
# of classrooms in schools type 1 in the district (divided by 100)	-0.995	-1.038	-0.978	-0.933
	(0.286)***	(0.283)***	(0.249)***	(0.251)***
# of classrooms in schools type 2 in the district (divided by 100)	-0.093	-0.088	-0.180	-0.203
	(0.068)	(0.069)	(0.070)***	(0.068)***
# of classrooms in schools type 3 in the district (divided by 100)	0.127	0.125	0.165	0.169
	(0.039)***	(0.040)**	(0.041)***	(0.039)***
% of pop. w/o sewage in the district	-0.243	-0.220	-0.004	0.004
	(0.283)	(0.285)	(0.276)	(0.279)
% of pop. w/o electricity in the district	-0.455	-0.446	-0.382	-0.403
	(0.243)*	(0.244)*	(0.244)	(0.245)*

Coefficients from ordered probit regressions reported, marginal coefficients for each category are available upon request. Robust standard errors clustered at the school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Source: ENRE 2001, School census 2002, Basic Statistics 2002.