Ibero-Amerika Institut für Wirtschaftsforschung
Instituto Ibero-Americano de Investigaciones Económicas
Ibero-America Institute for Economic Research
(IAI)

Georg-August-Universität Göttingen
(founded in 1737)

Diskussionsbeiträge · Documentos de Trabajo · Discussion Papers

Nr. 198

**Determinants of Student Achievements in the
Primary Education of Paraguay**

Thomas Otter, Carlos Villalobos Barría

December 2009

Determinants of Student Achievements in the Primary Education of Paraguay[*]

Thomas Otter & Carlos Villalobos Barría[†]

December 2009

Abstract

The idea that schooling scores depend on a combination of family background characteristics, ability and school (institutional) variables is quite clear. Regarding the issue of intergenerational transmission of inequality in the educational system, the most important question would be if and to what extent could a better institutional performance of the school service compensate for problems related to family background.

By means of the estimation of a reduced form equation for selected scores, we investigate the impact of institutional performance on scores after controlling for family background and individual characteristics. We do this by using a novel data set and an OLS and quantile regression approach to analyze how heterogeneous the process of score generation can be.

By providing integral health solutions, minimizing under-nutrition and providing ideal conditions in the classroom, training teachers can impact positively on low and mean learning outcomes, thus contributing to an improved educational quality and breaking cycles of intergenerational transmission of inequality. Increasing learning outcomes for levels above the median, only strengthens the transmission of inequality. Consequently, the equality approach should focus on trying to improve the worst scores and our results show that this can be reached at a significant level closing teacher training gaps, improving classroom conditions and improving health and nutrition.

1 Introduction

This paper addresses the question: why and how does an education system fail to provide its students with quality education? To find the answers, schools are one of the first places to search. In Paraguay, education quality is not constrained by the amount of expenditure because the country is spending a larger share of GDP than other Latin American countries and teacher salary levels or expenditure per student are also above the Latin American average. However, the qualifications of teachers, their performance in classroom, the quality of textbooks and materials and the motivational aspects of pupils seem to make up a recipe which leads to poor educational quality. We propose to estimate the impact of the different vectors on schooling achievements (Glewwe and Kremer, 2005) as a whole and in a second approach by means of quantile estimations (quantiles of levels of achievements). Quantile regressions will indicate whether each explanatory variable changes along the score's distribution or not. Learning achievement indexes can be ranked by school stratification. By doing so, we are able to analyze to what extent the differentiation by schools and by socio-economic sectors contribute to score's inequality. To achieve this, we use an inequality measure, traditionally used to measure income inequality. Regarding inequality, we find that improving the impact of our selected variables would have a positive effect on scores and additionally reduce inequality if this occurs for low income quantiles, but not for higher ones.

2 Education in Paraguay – an overview

Primary education lays the foundations for a wide range of competencies, mainly comprehensive reading, basic mathematical reasoning, the ability to do homework, or to work as part of a group. Students that do not master these abilities get stuck at a primary level, or even if enrolled at a secondary level, cannot move forward and eventually drop out from high school. Students from poor family backgrounds are especially affected by these limitations. For example, average national rates of repetition are estimated between 17 and 27% for first grade, but are almost double (30 to 50%) for the poorer half of the students' population (Schiefelbein and Brunstein, 2003). Furthermore, these high repetition levels (and their associated low achievement) have been almost constant since at least the 1980s (Zea-Barriga et al., 1981, 37; Schiefelbein and McGinn, 1980).

Results both from the SNEPE testing system and the UNESCO Adult literacy survey confirm low achievement levels in primary education (which generates high repetition rates). SNEPE

scores show that less than half of Grade 3 students can understand a brief text (less than 50 words) and that, eventually, only a third of the Grade 6 students will understand a 100-word paragraph in a front-page article of a national newspaper (Schiefelbein and Brunstein, 2003). In each case rural children (from families who are in the lower half of the income distribution) are achieving below the national average. Besides, the UNESCO Adult literacy survey found that only two thirds of the Paraguayan urban population sampled between 15-34 years old selected five or more right answers in a 7-item test rated as easy to answer ("literal" identifications). This score suggests that only half of the sample knew the correct answer for at least 5 of the 7 items (net score near 50%), the balance being only the result of random success (Schiefelbein and Brunstein, 2003).

Low achievement levels in primary education in Paraguay may be linked to a lack of access to pre-school for children from poor family backgrounds. "Investment in early childhood education is of key importance for building a strong foundation for lifelong learning and to ensure equitable access to learning opportunities later in school" (OECD, 2002, 182). Paraguay has a high "net pre-school enrolment rate", but there may be few opportunities for children from low income families to attend pre-school. This is a topic to be explored in the future with data from the Household Survey as a possible cause of low achievement in primary education. Lack of resources is another possible cause; therefore, data on expenditure is analyzed in the next section to verify whether resources are an effective constraint to delivering education of the required quality.

Paraguay provides universal access to the first grade of primary school[1] and keeps the newcomers enrolled for six or more years, but only 64% of each cohort finish primary education because teaching low-income students is ineffective and many students repeat grades.[2]

Almost 97% of children aged seven are "timely" enrolled. Dropping out begins between the ages of 9 and 11 when 10% of the cohort leaves school. Dropping out at the age of 11 or younger corresponds mainly to students that need personalized attention to learn.[3] The im-

---

[1]This figure is sometimes questioned, but Household Survey information on access is 98.3% while the Ministry of Education reports 99.4%; therefore, the difference is (at most) 1.1 percentage points.

[2]The analysis is focused on the "access to first grade of primary school" because the high coverage now offered in Paraguay reduces the reliability of usual indicators. For example, the "literacy rate" (self-assessment reported to the Census interviewer) is near 100% and the "Net Enrollment Rate" (for the 7-12 age group) is an ambiguous average of a distribution of single age rates (the range limits for "Enrollment Rates by single ages" in ages 7 to 12 are 89.3 to 99.4%). The graduation rate from sixth grade has an upper limit near 64% (some 60% enter secondary education according to the Household Survey data).

[3]About 10% of the population needs personalized attention (and facilities) for learning. Most Paraguayan schools cannot offer such special care. Therefore, students that need special attention either do not enroll or drop out early. There are economic variables linked to those earlier drop-outs involving special needs. In a richer coun-

Table 1: Enrollment rates and school expectancy 2000

Country	Years and range of high enrrollment[1]		Net enrrollmet rates by age			School Expectancy		
			<4	<5<14	<15<19	Primary and lower Sec.	Upper Sec.	Tertiary
Paraguay	5 years	G. 7 - 11	48.4	98	57.3	9.2	1.4	0.6
Brazil	8 years	G. 7 - 14	24.6	90.1	78	10.9	2.6	0.9
Chile	9 years	G. 6 - 14	23.6	92.7	66.7	8.4	3.5	1.7
Mexico	7 years	G. 6 - 12	35.5	94.8	41	9.4	1.4	1
Uruguay*	9 years	G. 6 - 14	23.5	97.8	60.7	9.9	2.4	1.8
Average OECD	12 years	G. 4 - 15	63.8	97.9	77.3	9.4	3.6	2.5

[1] At which over 90% of the population are enrolled.
Note: Lower secondary usually consists of three to six years of schooling.
Upper secondary could be terminal (preparing for a direct entry into working life) or preparatory.
School expectancy must be compared with the length of "primary and lower secondary education" in each country.
Sources: OECD, Education at a Glance, 2002. OECD Indicators 2002, pp. 220-1

pact of economic pressures begins at the age of 13 when 12% of the age group drops out (the difference between 89.3 and 77.7% in the last column). Dropping out accounts for 9% of the group aged 14 and similar rates are observed for the subsequent age groups. About 90% of the students remained enrolled in primary school for at least six years. However, only 60% of each population cohort enrolls in first year of secondary education. This attrition is linked to family income and education levels and poor reading comprehension that it is associated with repetition. Both causes may also be linked to not attending preschool education.

Enrolment data show that primary education coverage increased by about five percentage points relative to the previous decade (UNESCO-OREALC, 2001, 316) even though enrolment data could have an error of up to 3% (because enrolments may be over-reported or age dispersal may be underestimated). However, figures from different sources are consistent enough. The 2001 Household Survey showed that only 1.7% of the group aged between 10 15 years did not enroll in school (DGEEC, 2001) and therefore at least 98.3% had access to grade one in primary education (only one percentage point below the highest figure in the last column of Table 2). Hence, the probable level of error in enrolment data does not significantly affect the analysis of students' flows (carried out by comparing the net enrolment rates for each single age) presented below.

try an equivalent student would not end up dropping out partly because in some cases parents would pay for special attention, in other cases the schools would have the economic and human resources needed to deal with such cases, and in part because both parents and teachers would identify the problem better and earlier (in those cases where what is needed is not so much additional special resources as simply a proper diagnosis). Depending on the magnitude of the problem, the unit cost per student (to provide them a fair instruction) may be three to ten times higher than the current average unit costs.

Teachers make a student repeat Grade 1 when the student cannot decode simple words. Most Grade 1 students who are aged eight or older are repeating Grade 1 because there was timely enrolment of 96.5% of the group aged seven years (Table 2). At most 5070 newcomers who enrolled in 1999 in Grade 1 at the age of 8 ([100.0-96.5%] multiplied by 143,843) and 869 who enrolled in 1999 in Grade 1 at the age of 9 (0.6% multiplied by 141,694) should not be counted as repeaters.[4] However, some 10% of the students aged 6 that enrolled in 1998 (some 98,000, because the figure was 99,237 in 1999) may be repeaters in 1999. For the sake of simplicity both "corrections" have not been included in the "low estimate" of first grade repetition presented in the next paragraph.[5]

A "low estimate" of repetition in Grade 1 (data from 1999) corresponds to students of 8 years and older (given than 96.5% of the seven year olds enrolled on time). The 34,578 repeaters represent 17% of repeaters in Grade 1, but some have repeated several times. In fact, the total first grade repetition is 57,615 student-years (the number of "over-aged students" multiplied by the number of "extra years") and represents a high cost in resources (because most of those students should have learned to "decode a short written text" in one year, and be promoted to second grade). The "low estimate" of the repetition rate (Table 2, last row) slightly exceeds the repetition rates reported by the school principals (UNESCO-OREALC, 2001, 318), but is below the 20.2% failure in first grade reported in 1999 (PNUD, 2003, 172).

A "high estimate" of repetition in Grade 1 assumes that the share of newcomers aged between 5 and 8 years is near constant over time (according to distance to school, sex, talent distribution, or traditions) and that the percentages – for each of these four ages (regarding each population) – should add up close to 100% (newcomers should be in a range of 97.7 to 99.4 or 100% of the population of the normal entrance age according to the Household Survey and institutional data). Given that over 90% of newcomers are aged 6 or 7 years and the small difference in both populations (the difference between 145,878 and 143,843 being less than 2%) the 7 year old group is used as the basis.

The repetition figures for Paraguay are well above the levels in OECD countries, but they are similar to the averages in other Latin American countries (UNESCO, 2001, 4 and 44; Wolff et al., 2002). Moreover, there are differences in repetition levels according to family income levels

---

[4]Assuming that 100% of the cohort was enrolled and that (100-99.4%) dropped out after one or two years of schooling.

[5]There is a difference between the "number of repeaters in grade 1 of primary as of 1999" and the "total number of years repeated by students that are in grade 1 of primary in 1999". Most children with normal talent and behavior would not repeat grade 1 more than three or four years. However, parents with children with learning difficulties or requiring special assistance may keep them more years in first grade (because there may be no work opportunities for them or consider the school as a low-cost caring alternative).

as shown in the next paragraph.

The rate of repetition is usually low (about 4 or 5%) in schools for students in urban non-marginal areas, and rather high (between 30 and 50%) in urban-marginal and rural areas. If Paraguayan schools were split into two halves (according to family income), the national average would correspond to half of the sum of the repetition rates of the wealthier and poorer halves. For example, repetition would approach 30% in poorer areas if the rate were 4% in the wealthier areas – according to the "low estimate" for Paraguayan schools in 1999 – and the corresponding figures would be 50 and 4 for the "high estimate". These assumptions (for wealthier and deprived areas) – and the implication on the impact of teaching methods – suggest the need for estimating repetition by family income levels.

If most of the average students enrolled in Grade 1 are students from the lower half of the income distribution, their teachers will have a tough task teaching such heterogeneous classes. High repetition constrains the education levels of children from low-income families (few of them reach secondary education) and increases the age-heterogeneity in classrooms reducing the effectiveness of frontal teaching. Furthermore, repetition is also linked to dropping out (and to family income levels) as discussed in the rest of this section.

In general, dropping out is generated by several causes. It is probable that children with physical or mental limitations are forced to drop out first because the regular system does not deliver the special attention that they need. Most of these early drop-outs (aged 11 and 12) would correspond to students with special attention needs that are not available in the regular schools where they are enrolled, although only one out of ten (1% of the population) needs to be in special institutions outside the regular system of education (UN, 1987, 53). Dropping out generated by other reasons –mainly pressure to contribute to the family income– really begins around the ages of 13 and 14 years when some ten-percent of that age group drops out. This group of dropouts was enrolled for some eight years, although some dropped out temporarily during harvest periods. Paraguayan data suggest that these students did not complete primary education because they repeated several times due to poor achievement.

In summary, the school system has built enough space to take care of each new population cohort reaching school age, but substantial repetition linked to poor achievement and eventual dropping out distorts the enrollment pyramid reducing the access to higher grades. Repetition and dropping out (more recurrent in students from low-income families) generate a large gap between newcomer students to grade 1 in primary education and the subset starting secondary education. Grade repetition and early dropouts eventually reduce the number of children from low-income families enrolled in secondary or higher education. Repeaters and dropouts tend

to be members of low-income families.

Despite the significant progress Paraguay has achieved in recent years regarding access to education and school permanence, the country still faces great challenges. These include the problems of internal efficiency, quality in learning and equity. For instance, retention rates indicate that half of school children finish their 6th grade without having to repeat any grade (44% in rural areas and 62% in urban areas). Regarding children's learning, the low scores achieved in the SNEPE's tests of 2004 (3rd grade: 54% for language and literature and 58% for math; 6th grade: 60% for language and literature and 63% for math) show that the quantitative growth of the system is not ensuring an adequate level of learning.

In Paraguay education quality is not severely constrained by the amount of expenditure because the country spends a larger share of GDP than other Latin American countries and teachers' salary levels or expenditure per student are also over the Latin American average. However, the qualifications of teachers, their classroom performance, the quality of textbooks and materials and the motivational aspects of pupils seem to form a recipe which results in poor educational quality.

Total expenditures on education as a percentage of GDP in Paraguay are higher compared with other countries of the Southern Cone, Mexico and even the average (Table 2). The percentages for upper secondary are over twice as high than all other countries, and only expenditures in tertiary education are below the figures for Chile and for the average OECD country. However, given that GDP figures may be underestimated (see paper by N. Aguilera) this conclusion should be eventually revised. Furthermore, there is a balanced contribution of both public and private sources (Table 2). Funding for education is provided both by direct public expenditure (4.8% of GDP) and through private spending in educational institutions (3.7%). Paraguay spends on public education the same as the median spending of "countries with high human development", and above the 4.2% in "medium human development countries" where Paraguay is included (UNDP, 2003, 93). These figures show a comparatively high level of resources being allocated to education. However, it is necessary to confirm whether funds are used to pay teachers suitable salary levels and to assess the level of expenditure per student comparing it with the expenditure per student in other Latin American countries.

Mid-career salaries for Paraguayan teachers relative to GDP per capita are high compared with the countries of the Southern Cone and Mexico (Table 3). Teacher salaries is a key variable for attracting good candidates for initial teacher training -keeping trained teachers in the profession- and reducing the real school calendar for time spent on strike. A high salary should have a positive impact on these three aspects and should also improve students' achievement,

Table 2: Total expenditure on education as percentage of GDP (1999)

Country	Percentage of GDP by educational levels					For all levels of education	
	Pre-Upper Sec Secondary	Upper Secondary	Tertiary	Total[1]	Public	Private	Estimated Total[2]
Paraguay	4	2.8	1.5	8.5	4.8	3.7	8.5
Argentina	3.2	0.8	1.1	5.8	4.5	1.3	5.8
Brazil[3]	3.4	0.6	1.1	5.1	5.1	n.a.	6.2
Chile	3.6	1.4	2.2	7.2	4.1	3.1	7.2
Mexico	3.3	0.8	1.1	5.2	4.4	0.8	5.2
Uruguay[3]	1.9	0.5	0.6	3	2.9	n.a.	4
Average OECD	2.3	1.3	1.6	5.8	4.6	1.1	5.8

[1] This includes undistributed expenditure and advanced research programs.
[2] This includes estimated private expenditure similar to the average OECD.
[3] This includes only direct public expenditure by educational levels.
Sources: OECD, Education at a Glance, 2002; OECD, Indicators, 2002, pp. 170-72.

reduce repetition and raise the quality of Human Capital in Paraguay. On the other hand, teacher salaries make up near 90% of the cost of providing education in Paraguay, making this variable a crucial element in decision making in education. There is a trade-off between better-paid teachers and a balanced education budget.

"Salaries and working conditions of teachers, including starting salaries and pay scales, and the costs incurred by individuals in becoming teachers, compared to salaries and costs in other high skill occupations are key factors in determining the supply of qualified teachers" (OECD, 2002, 332). The comparison of teachers' salaries to GDP per capita provides an indicator of the relative value of teachers' salaries and also of affordability for countries (Table 3). Paraguayan teachers get 2.0 to 3.1 times GDP per capita. These ratios are better than Brazil (1.5 and 2.4), Chile (1.4), Mexico (1.6 and 2.1) or even Uruguay (1.6 assuming full-time workload). The Paraguayan ratio is also higher than the average OECD country (1.3). In fact, "the ratio of the average upper secondary teacher's salary to GDP per capita is the second of all WEI [World Education Indicators] countries at 3.13 and above all OECD countries" (OECD-WEI, 2003, 137).

In Paraguay, teacher salaries increase with the level of education (Table 3). There is no increment linked with years of experience, while such an increment is near 40% in Brazil, in Chile 10%, 30% in Mexico, 20% in Uruguay, 90% in New Zealand and 40% in the average OECD country. The salary in Secondary Education (with training) is 60% higher than the salary in Primary Education. This percentage is similar to the situation in Brazil (also 60%), and higher than in Mexico (30%) or the other countries presented in Table 3. In the average OECD country the salary in Secondary Education (with training) is only 10% higher than the salary in Primary

Table 3: Ratio of teacher salaries after 15 years of experience to GDP per capita, 2000[1]

Country	Salaries in Primary Education(with training)			Salaries in Secondary Education (with training)		
	Starting	15 years	Ratio to GDP/P	Starting	15 years	Ratio to GDP/P
Paraguay	8,874	8,874	2	13,911	13,911	3.1
Brazil	7,420	10,176	1.5	14,820	16,240	2.4
Chile	10,716	12,038	1.4	10,716	12,038	1.4
Mexico	11,235	14,824	1.6	14,383	18,760	2.1
Uruguay	5,749	6,891	0.8	5,749	6,891	0.8
New Zealand	17,354	33,653	1.7	17,354	33,653	1.7
OECD average	21,469	29,407	1.3	22,727	31,221	1.4

[1] Equivalent US dollars converted using purchasing power parities PPP.
Note: Salaries based on a 20 hour per week workload. Most teachers hold two positions. Figures in the last column should be further revised given that the GDP estimations are distorted for certain years. This includes estimated private expenditure similar to the average OECD.
Sources: OECD, Education at a Glance, 2002; OECD, Indicators, 2002, p. 339.

Education. Of course, salaries per teaching hour in Paraguay are also higher in secondary education, since secondary teachers (as in most countries) are required to teach fewer hours than primary teachers.

Even though the Paraguayan expenditure per primary student is the lowest, the country invests in each student a higher percentage of its GDP per capita than the rest of the countries of the Southern Cone and Mexico (Table 4). The analysis of the "expenditure per primary student" indicator is tricky because the short-term and long-term effects must be taken into account. For example, the expenditure per primary student depends on the salary paid to the teacher, but a sudden salary increment cannot change the training of the teacher. The same teacher will still be teaching tomorrow, in spite of receiving more money than today.

"The expenditure per student indicator shows direct public and private expenditure on educational institutions in relation to the number of full-time equivalent students enrolled in each level. ... The variation in expenditure on education per student may reflect not only variation in the material resources provided to students (e.g., variations in the ratio of students to teaching staff) but also variation in relative salary levels" (OECD, 2002, 147).

Teacher salaries have a major impact in overall unit costs given the traditionally labour-intensive teaching processes. Increments in salaries have a proportional increment in the expenditure per student. Student-teacher ratios also have an important effect; Paraguay had 21 students per primary teacher in 1996 (UNESCO, 2001, 620), therefore an increment up to a 30 student-teacher ratio would reduce the "annual expenditure per student" (a proxy of the unit cost) by 30%. On the other hand, a drop in enrolments may lead to a significant increase in spending

Table 4: Annual expenditure per student by level and as a percentage of GDP per capita (1999)[1]

Country	Expenditure per student by level of education			Expenditure as a percentage of GDP per capita		
	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary
Paraguay	877	1,545	5,465	20	35	125
Argentina	1,629	2,327	5,606	13	19	46
Brazil	956	1,100	13,567	14	16	195
Chile	1,701	1,941	6,911	20	22	80
Mexico	1,096	1,480	4,789	13	18	57
Uruguay	1,000	1,275	2,239	11	14	25
Average OECD	4,148	5,465	9,210	19	25	44

[1]Equivalent US dollars converted using purchasing power parities PPP.
Sources: OECD, Education at a Glance, 2002; OECD, Indicators, 2002, pp. 158-59.

on education per student (OECD, 2002, 150).

The annual expenditure per student in Paraguay increases significantly with the level of education as it happens throughout the OECD countries (see the OECD averages in the last row in Table 15). At the secondary education level, the annual expenditure per student is 1.8 times the expenditure per student at the primary education level (80% higher). This increment is higher than in other countries of the Southern Cone (1.1 to 1.4), Mexico (1.4) and the OECD average (1.3). At the tertiary level, the annual expenditure per student is 6.2 times the expenditure per student at the primary education level. This increment is lower than the increment in Brazil (14.2), but higher than the rest of the countries of the Southern Cone (ranging from 2.2 to 4.1), Mexico (4.4) and the OECD average (2.2).

3   The Data

We use three different data sets, namely the Paraguayan household survey (Encuesta Permanente de Hogares – EPH 2005), the World Education Indicators' Survey of Primary Schools (WEI-SPS), prepared by UNCESO (data from 2006) and data from the SNEPE (Sistema Nacional de Evaluación del Proceso Educativo) survey in Paraguay in 2003. The WEI-SPS survey was applied to students in grade 4 and SNEPE exams were applied to students in grade 3 as well as to teachers in training (future teachers). The idea of this research was to use the learning outcomes score variable provided by the SNEPE data base as the dependent variable and build up a combined data set with information from SNEPE, EPH and WEI-SPS for the explanatory variables.

The SNEPE data base provides information on the results of mathematics and reading / writing tests as a percentage of correct answers and as the associated qualification in school grades. We used the mathematics and communication scores in its pure form (non-logarithmic) as dependent variables. In general, the way our model is specified, it can be understood as a reduced form specification where the coefficients change in their scope according to the specification of the production function.

4   Previous Research

Since the mid-1960s, a vast amount of literature has unfolded in the United States and Europe in an attempt to estimate some production function of learning. Hanushek (2002) accounts for 376 published estimates of US education production functions from 89 separate articles published in peer reviewed journals or books until 1994.

Apart from the US and European evidence, there are also several studies on education production functions in developing countries, summarized e.g. in Hanushek (1995) and Glewwe (2002). While many of the older studies for developing countries have been criticized for their lack of methodological and data quality, several more recent studies have presented convincing evidence on education production functions for countries in Latin America, Africa, and Asia (see the references in Glewwe 2002). In particular, several studies use data from randomized trials to estimate the impact of distinct educational policies on student performance, most recently in Kenya (see the overview by Kremer 2003), and quasi-experimental evidence on class-size effects exists for South Africa (Case and Deaton 1999), as well as Israel (Angrist and Lavy 1999). While nearly all of these studies are confined to individual countries, recent internationally comparable evidence exists for several countries in francophone sub-Saharan Africa (Michaelowa 2001) and for several newly industrializing countries in East Asia (Wößmann 2003a).

5   Approach and methodology

Glewwe and Kremer (2005), propose to estimate the production function for learning that may lead to future earnings according to a structural relationship that can be depicted as:

$$A = a(S, Q, C, H, I) \tag{1}$$

where A represents the skills learned (achievement), S represents the years of schooling, Q represents a vector of school and teacher characteristics (quality), C represents a vector of child characteristics (including "innate ability"), H represents a vector of household characteristics, and I represents a vector of school inputs under the control of parents, such as children's daily attendance and purchase of textbooks and other school supplies (endogenous vector).

Based on the data available in Paraguay, primarily a national survey on educational achievements for grade 4 (see below), we propose to estimate the reduced form relationship.

$$A = k(C, H, L, EP) \tag{2}$$

Where all vectors are exogenous, L represents the local community characteristics and EP the education policies. L and EP may interact to determine the quality (Q) of a school and even the prices (P) of educational inputs such as fees.

$$Q = q(L, EP) \tag{3}$$

$$P = p(L, P) \tag{4}$$

In principle, a reduced form equation can be estimated as A=h(Q,C,H,P), but because of the impossibility of getting all prices (P) and schools characteristic which determine quality (Q) and given our available data, we abandoned attempts to estimate the reduced form and we estimate the above reduced form relationship which directly links education policies and local characteristics to our available educational outputs.

In our case, achievements and skills can be operationalized throughout scores of standardized tests for the areas of reading / writing and mathematics.

We propose to estimate the impact of the different vectors on schooling achievements as a whole and in a second approach via quantile estimations (quantiles of levels of achievements). Quantile regressions will indicate whether each explanatory variable's effect changes along the score's distribution or not.

Additionally, we are able to analyze to what extent differentiation by schools and by socio-economic sectors contribute to score's inequality.

To achieve this, we use an inequality measure, traditionally used to measure income inequality: the Theil 2 Index. The Theil 2 measure -the mean logarithmic deviation measure- is defined

as:

$$T_2 = \frac{1}{N}\sum_{i=1}^{N} ln\left(\frac{\mu}{A_i}\right)$$

where A_i represents the student i test score, N represents the total number of students and μ represents the mean test score of the distribution. This index is zero for the case of perfect equality, approaches infinity in the case of perfect inequality, and can take intermediate values between the two.

We choose this inequality index because it satisfies a set of axioms and properties that are convenient when measuring inequality: *symmetry* (the measure is unchanged if there is a permutation of test scores between two persons; this principle is also called the *anonymity* principle); *replication invariance* (the measure is unchanged if the population is doubled, tripled, and so forth), *mean independence* (the measure is unchanged if all scores in the distribution are multiplied by a scalar); and the *Pigou-Dalton Principle* (the inequality measure increases with any regressive shift in the (same total) scores from low to high performers. The Gini Index also satisfies these properties, so we will report results of both indexes further ahead.

5.1   Quantile approach – Median regression

Consider the following distribution function of a random variable Y

$$F(y) = Pr(Y \leq y) \tag{5}$$

For any $\tau \in (0,1)$, the quantile of Y is defined as:

$$Q(\tau) = \inf\{y : F(y) \geq \tau\} \tag{6}$$

The quantile function gives a complete characterization of Y. $Q(0.5)$ is the median, the first quartile is $Q(0.25)$, the first decile is $Q(0.1)$ and so on. We can now write (3) in a different way. Let $x_i, i = 1, ..., n$ a $(K \times 1)$ vector of explanatory variables, then:

$$F_{u\tau}(\tau - x_i^{'}\beta_\tau|x_i) = Pr(y_i \leq \tau|x_i) \tag{7}$$

This equation is equivalent to

$$Y_i = x_i' \beta_\tau + u_{\tau i} \tag{8}$$

where the only constraint is that $Q_\tau(u_{\tau i}|x_i) = 0$ and the distribution of this error term left unspecified. Now we can write the linear conditional quantile function as

$$Q_\tau(\tau|X = x) = x_i' \beta_\tau \tag{9}$$

Where β_τ can be estimated solving

$$\hat{\beta}_\tau \quad = \quad \arg \min_{\beta \in R^K} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta)$$

Where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the piecewise linear "check function" and $I(.)$ is the indicator function. The quantile regression estimator achieves a more complete description of the conditional distribution of Y given $X = x$. The partial derivate of the conditional quantile of y (see(7)) with respect to one of the regressors, $j - th$, could be interpreted as the marginal change in the $\tau - th$ quantile due to a marginal change in the $j - th$ regressor. Note that, this marginal effect is related with τ and not with some particular individual, which changes their $\tau - th$ quantile simultaneously when some value of their explanatory variable changes.

There is a mayor advantage using Quantile regression. We can detect and deal with the plausible heteroskedasticity in the data allowing different slopes for each conditional quantile in the score distribution. Using the method suggested by Koenker and Bassett (1978, 1982) to obtain the standard errors, Rogers (1992) reports that these standard errors are suitable in the homoskedastic case but that they look to be understated in the presence of heteroskedastic errors. Accepting the heteroskedastic nature of our data, we proceed to bootstrap[6] the estimated coefficients in order to obtain the standard errors relaxing the homoskedastic assumption and improving our inference power.

The median regression is also more robust than an OLS regression in the presence of outliers and in contrast to the Heckman ML estimators, it does not rely on the normality of the residuals

---

[6]The bootstrap is a computational intensive procedure for assessing uncertainty in estimation, in which through resampling data replaces mathematical analysis. We use the bootstrap to attach a standard error of the estimated parameter.

because the estimation is only influenced by the local behavior if the conditional distribution near the specified quantile.

5.2 Some comments on applied econometrics

The first estimation method applied in this paper is a linear OLS regression for survey data. This estimation technique relies on the normal distribution of residuals with zero mean. The fact that unobserved characteristics could be correlated with some explanatory variable potentially introduces a bias in our estimates. In order to avoid this source of bias, we exploit our data base in order to minimize the possibility of observed characteristics remaining in our residual, and in this way reducing the probability of an error of specification (omitted variable problem).

Although extensive reviews of related literature exist[7], much controversy remains over what conventional estimates of education production functions can and do tell about the causal relation of student performance, for example with respect to the effectiveness of class-size (Hanushek, 2003 and Krueger ,2003).

The main data base used in our work is SNEPE, which identifies cluster (school id) and strata (public-urban, private-urban, public-rural and private-rural), and therefore we adopt the sample design in order to optimize the precision of our estimators. In order to incorporate other indicators, we merge our main data base with the mean values for each of interest variable at the district level, provided from WEI-SPS and EPH data bases.

This methodology is more powerful than the usual (frequently cited in literature) weighted OLS regression approach, given that it controls for cluster correlations. Control for cluster correlations in our case basically means that we take into account the fact that in certain sectors we might have several schools that are quite homogeneous. In this case, variation will not be a pure (100%) random effect. Ignoring these cluster effects would imply considering each observation as independent, which would create artificially higher variation coefficients, a under estimation of standard errors and possibly accepting statistically significant results when in reality thay do not exist. Standard errors in our estimation are robust because they are clustered at the school level.

For quantile regressions we used the bootstrapping procedure considering the clustered struc-

---

[7] About the specification and interpretation of education production functions, Hanushek (1986; 2002) offers an excellent discussion.

ture of our data and ran these with R software[8]. This way we basically get almost the same statistical rigor in quantile regressions as in OLS (via bootstrapping with 500 replications[9]).

The bootstrapping procedure which takes into account the clustered design of our database, additionally allows not to make assumption about the underlying score distributions and is therefore a proper way to relax those exigent distributional assumptions concerning the nature of the residual structure in the regression models[10].

6 Results

Tables 1 and 2 in the appendix show descriptive statistics of our joint data base of EPH, WEI and SNEPE data, expressed as quintile means. Note that since we have no income data in our data base, quintiles refer to test scores. (T); this is true for the descriptive statistics below as well as for the quantile regressions farther below in the text.

We have divided our results into four different groups: students' personal characteristics (C) such as age, gender, motivation indicator; family background (F) such as parental education, language, coeducational vocation of parents; Local Characteristics (L) such as district average child sickness rate, district average health insurance rate, district undernourishment rate and urban/rural public/private status; and Educational Policies Variables (EP) such as school language, optimal classroom conditions and school size, which are treated as exogenous. Since there is no self-selection into the sample as 3rd grade schooling is compulsory[11], least-squares estimates should represent causal family-background effects. It helps to clarify in advance what the estimates of the coefficients on the family-background variables stand for. Given the technical constraints on the pedagogical process, the size of the effect of any family-background characteristic on students' educational performance should be the same everywhere. If this is not the case, this implies that there must be differences in how the school systems work.

According to Woessmann (2003), As to the main family-background effects we have the education level attained by the parents, which is strongly related to student performance; the

---

[8]For our estimation we use the quantreg package developed by Roger Koenker.

[9]Efron and Tibshirani (1993) demonstrates that 200 replication provides enough information to properly compute the standard deviations of our estimates. We run 500 replications considering that the object oriented R software allows to do this with extremely low cost in time.

[10]This is important because our data set contains variables at different aggregation levels.

[11]According to data from EPH 2005 the gross enrolment rate for boys and girls aged 8 (official age for 3rd grade) is 97.8% (self-processed data).

social and economic background of the students in general and families unobserved aspects of family background, such as parent's motivation or their willingness and capability to help with homework. In general, it seems to be that a larger family-background effect materializes in terms of better educational performance of the children rather than in terms of higher attainment levels.

Additionally, least-square regression results of student performance on schools' resource endowments and of teacher characteristics, controlling for family-background influences, might suffer from endogeneity bias. The resource endowments that students face may be particularly prone to endogeneity, in that resources and student performance are simultaneously determined by other factors or that the latter is even directly used to determine a reverse causality (Hoxby 2000). For example, school systems may sort weaker students into smaller classes, giving rise to a negative correlation between performance and the teacher-student ratio in classes (West and Wößmann 2003). Also, parents of high performing students may choose to move into areas where schools are better resourced, giving rise to a positive performance-resource correlation. None of these correlations say anything about the causal effect of resource endowments on student performance. In short, parents, teachers, schools, and administrators all make choices that may give rise to a non-causal relationship between resources and student performance, so that least-squares estimates of resource effects may be biased.[12] Of course, there are some possibilities to control for this effect. For example, some recent influential studies have tried to avoid endogeneity biases in the class-size estimates by using experimental evidence from the Tennessee Project STAR (Krueger 1999) and quasi-experimental research design for data from Connecticut (Hoxby 2000); one of them finding substantial class-size effects while the other finding none at all. Nevertheless, in practice it is almost impossible to control for all possible biases, mainly due to limitations in the data itself.

Tables 6 and 7 in the appendix show results for OLS full specifications. Tables 8 and 9 in the appendix show results for the same specifications estimated by means of quantile regressions technique.

6.1 Language *(F, EP)*

Surprisingly, students from Guaraní speaking families in our data show a higher performance than those from Spanish or other language speaking households. By looking at the tables (ols

---

[12]Hoxby (2000) shows that simple least-squares estimates of resource effects are biased upwards in the case of class-size effects in the US state of Connecticut; i.e., the least-squares estimates are biased in the direction of finding positive effects of smaller classes on student performance.

math, ols com, quant math, quant com), it is clear that the family language effect on communication score concentrates towards the upper tail of the distribution, while for Mathematic scores the effect appears to be more equally distributed and of greater magnitude. This result, contrary to the expected outcomes and inconsistent with other empirical evidence[13], could be produced by a problem of selective sample design, where schools with a higher native or indigenous background were over-represented, and thus sub-representing at the same time (elite or higher qualified) schools of non-native urban students.[14] Nevertheless, Spanish-speaking teachers on average have higher performing students (ceteris paribus) than teachers who speak any other language.

The impact of the language usually spoken in the household is smaller than the impact of teacher's language. This way we observe that teachers who are teaching in languages other than Guaraní have a negative impact on student performance compared to teachers teaching in Spanish.

The institutional language variable (language spoken by the teacher) is powerful and controls for the language spoken in the household. This way we observe a pure association (correlation and non-causality proved) between the teacher's language and the student's performance. Quantile regressions and OLS show for communication (Spanish) and mathematics a robust impact in all specifications. Our findings could indicate a "low teaching quality" of "native", mostly rural or sub-urban Guaraní speaking teachers. A differentiated teacher's training for non-Spanish speaking teachers might help to improve this situation.

The effect on scores follows an inverted U-shape, where students who are concentrated precisely around the median obtain approximately one score point more than students who are taught in a language other than Spanish. These issues therefore limit the progress of good students, while also precluding the improvement of poorly performing students.

---

[13]For instance, the JICA Report on the Paraguayan education sector (JICA, Instituto Desarrollo 2004) mentions a double limitation which arises from the dilemma of a population of which at least 75% speak Guaraní as mother tongue but implements (as a teacher) and attends (as a student) a Spanish speaking schooling system. Many pupils do not speak Spanish adequately when entering school and have difficulties keeping pace with the classes, meanwhile teachers whose mother tongue is Guaraní frequently show limited capacities in adequately transmitting curricular contents in Spanish.

[14]The number of well-performing urban schools with Spanish speaking teachers is in fact lower than the number of poorly-performing rural, sub-urban and even urban schools. If sub-representation of these generally better performing schools is so low that we get only a very limited number of cases, then there is space for a random selection of the worst performing Spanish speaking urban schools, which could create this kind of bias.

6.2 Under-nutrition *(L)*

Under-nutrition is defined as the mean district under-nutrition rate and belongs to the local community characteristics (vector L). This variable shows the expected negative outcomes on schooling performance. Given its nature, it can be considered as exogenous from the point of view of households, only and only if we suppose that there are local factors that determine the district rates of under-nutrition an the nutrition of each of the children in the district.

Our results clearly show that higher under-nutrition rates impact negatively on schooling performance concerning the median and below. This variable has a U-shaped impact on scores. Since the nutrition variable is a district mean, and not individual information, it does not clearly show that nutritional problems might explain poor schooling performance on the average. Nevertheless, it shows that good school performance is immune to nutritional problems basically because we can suppose that good performing students do not suffer problems of under-nutrition. The overall nutritional impact is quite important. School performance of districts with zero under-nutrition is 3 to 6 points higher than the districts with the highest levels of nutritional problems. It is important to keep in mind that the nutritional variable ranks from 0 to 1. We can therefore roughly conclude, for example, that reducing average under-nutrition from 0.5 to 0.4 could improve schooling performance between 0.3 and 0.6 points (according to our data).

Consequently, our data show that poor parents are likely to produce poorly nourished children with low schooling performance, and therefore low human capital accumulation, resulting later on in probable low incomes and ongoing poverty. But given that there also exist spatial patterns of poverty (poorer areas or poverty pockets within better-off areas), under nutrition can simultaneously have endogenous and exogenous characteristics.

Note that in our data the impact of nutritional problems is higher in mathematics than in communication.

6.3 Optimal class room conditions *(EP)*

This is a constructed variable that includes information provided by the students on the existence of chairs, desks, electricity and the individual feeling of being comfortable in the classroom. Even if this variable seems to have a subjective background, it is objective in the sense that students who do not feel comfortable will not express the opposite feeling. The variable

is operationalized as a dummy variable[15] and is robust in OLS and quantiles,[16] with impacts stronger on mathematics than on communication and homogeneous impacts throughout the distribution, which basically reflects that both good and bad students benefit in a similar way from good classroom conditions. Additionally, if students have varied backgrounds, optimal conditions would be a progressive factor, given that they offer an environment of equality in the classroom for students coming from unequal homes.

6.4   Small-scale school *(EP)*

We have to be quite careful with the interpretation of this variable, given its problems of endogeneity. Small schools are defined following a classification adopted by the Paraguayan educational authority[17]. Small schools can impact the score, but at the same time the score can impact on the school's scale. For example, a small school can in fact benefit the median, but at the same time the scale can be a consequence of better families who select small schools for their children in order to get personalized teaching in small groups; this way a higher score can determine the school's scale. Within our data set the school size variable is most affected by endogeneity, meanwhile the literature cites school scale and class size as the most affected variables.

Thus we discuss their estimates only in terms of correlations. Small schools are associated, on the average, with higher performance. Quantiles show an inverse U shape indicating that students performing at a mean level benefit from attending a small school (according to the official classification from the Ministry of Education). Behind this observation, we might have the fact that small schools offer a less competitive or more familiar environment, which might be a limitation for better performing students or might be the consequence of the existence of small schools for students with whatever kind of inability or small schools as a response to the lack of resources. Whatever the real cause of the inverse U shape may be, small schools seem to be ideal for median performing students. Even if we do not have a strong evidence of causality for this, we can at least state that small schools would do no harm and offer the possibility to create an environment where other variables, positively related to schooling scores, could start to have a positive additional impact.

---

[15]Defined as 1 if the student reports to have adequate chair, desk and ventilation and that he additionally feels comfortable. Only if all four conditions are fulfilled is the dummy 1, otherwise 0 is assigned.

[16]We tried different possibilities for implementing this variable (tossing in three dummies instead of one constructed with four variables).

[17]For details see http://www.mec.gov.py/cmsmec/

6.5 District mean level of health insurance *(L)*

Our evidence seems to indicate that bad students benefit most from higher district levels of health insurance. We cannot, however, understand our quantile estimate results in a way that once we increase the health insurance level at a given district students will perform better. We rather should think about it in a way that health insurance rates are part of a more favorable learning environment and contribute to a better performance. So it is not that higher insurance rates make poor performance as bad as it is, but that without the positive impact of higher insurance rates, poor performance rates would be even worse.

6.6 Inequality analysis

Estimating inequality of scores and comparing them with the variation of regression coefficients by income quantile allows us to understand to what extent differentiation by schools and by socio-economic sectors contributes to score's inequality.

Table 5: Inequality indicators and regression coefficients for mathematic scores

Index-Coefficient	Q1	Q2	Q3	Q4	Q5
Theil 1	0.04408	0.00659	0.00278	0.00126	0.00125
Theil 2	0.03957	0.00656	0.00277	0.00126	0.00126
Gini	0.15702	0.06499	0.04197	0.028	0.02811
Spanish speaking teacher	0.80575	0.8656	0.86778	0.88219	0.89165
Classroom conditions	0.84633	0.8825	0.8958	0.90953	0.94688
Small size school	0.17021	0.16318	0.19282	0.26735	0.28568

Source: Own estimates based on joint database EPH 2005 (DGEEC), SNEPE 2003 (MEC)
WEI-SPS 2006 (UNESCO).

Table 6: Inequality indicators and regression coefficients for communication scores

Index-Coefficient	Q1	Q2	Q3	Q4	Q5
Theil 1	0.03763	0.00459	0.00264	0.00178	0.00415
Theil 2	0.03372	0.00457	0.00262	0.00178	0.00421
Gini	0.1431	0.05307	0.04004	0.03306	0.0513
Spanish speaking teacher	0.795754	0.859358	0.8703	0.881279	0.903749
Classroom conditions	0.842994	0.885979	0.898133	0.916362	0.937626
Small size school	0.203385	0.204609	0.201239	0.253325	0.270199

Source: Own estimates based on joint database EPH 2005 (DGEEC), SNEPE 2003 (MEC)
WEI-SPS 2006 (UNESCO).

Tables 5 and 6 show that, within lower score quintiles, inequality is generally higher for any inequality indicator we use and inequality generally decreases towards higher quintiles. This is entirely true for mathematic scores and is true for quintile 1 to 4 regarding communication scores. Nevertheless, we observe an inequality increase for communication scores in the highest quintile. At the same time, regression coefficients of our three selected institutional (school) variables are higher for higher quintiles, so they have a bigger impact to increase score results. If score inequality is to be reduced most efficiently, considering the complete distribution, increasing scores in lower quintiles, we clearly find the opposite situation here. In a way, learning outcomes even seem to be "regressive". So even if it's true that improving the impact of our selected variables would have a positive effect on scores, we would only reduce inequality if this happens for low income quantiles, but not for the higher ones. If we improve the impact of our institutional variables throughout the complete distribution, inequality might even increase or at least remain the same.

6.7 Test of Robustness

Our reduced form assumes exogeneity of our explanatory variables. Table 4 and 5 in the appendix show the results for the mean regressions for the determinants of Communication and Math scores based only on household and student characteristics (vectors C and H). By comparing this mean regression with ones those who include additional explanatory variables and with the median quantile regression, it is clear that vectors C and H are well defined and appear to be robust in all specifications.

7 Conclusions

The idea that schooling scores depend on a mixture of family background characteristics, ability and school (institutional) variables is quite clear. Regarding the issue of intergenerational transmission of inequality in the educational system, the most important question would be if and to what extent a better institutional performance of the school service could compensate for problems related to family background.

By providing integral health solutions, minimizing under-nutrition and providing good conditions in the classroom, training teachers (according to their gaps) can impact positively on low and mean learning outcomes, thus contributing to an improved educational quality and breaking cycles of intergenerational transmission of inequality. Increasing learning outcomes for levels above the median, only strengthens the transmission of inequality. Consequently, the equality approach should focus on trying to improve the worst scores and our results show that this can be reached at a significant level closing teacher training gaps, improving classroom conditions and improving health and nutrition.

Score inequality can be reduced most efficiently, considering the complete distribution, by increasing scores in lower quintiles. So even if it is true that improving the impact of our selected variables would have a positive effect on scores, we would only reduce inequality if this occurs for low income quantiles, but not for higher ones. If we improve the impact of our institutional variables throughout the complete distribution, inequality might even increase or at least remain the same.

To ask education policies to focus their efforts regarding inequality-reducing measures on lower income quintiles should not be that difficult, given that an important part of students from higher income quantiles are enrolled in private education. Any action taken by public policies to reduce inequality will have a limited impact regarding its possibility to reduce the intergenerational transmission of (welfare) inequality, since family background is a much stronger force in determining inequality in welfare levels. Nevertheless, since parents education levels are a much stronger determinant regarding levels of welfare inequality, better performing education policies will not have a short term impact. But they can have an important long term impact, not only improving test scores but helping students to reach a higher level of education.

8 References

Alfonzo, Leonardo, 2000, La educación en la determinación de los ingresos salariales en el Paraguay, en Economía & Sociedad, N° 1, Asunción, pp. 29-48.

Barro, Robert J and Jong-Wha Lee, 2000, International Data on Educational Attainment: Updates and Implications, Centre for International Development Working Paper No. 42 April 2000, Harvard University Cambridge MA.

Behrman, Jere, 2003, Educational Sector Study: Pro-Poor Economic Growth Effects of Policies and Activities, Population Studies Center, University of Pennsylvania, April 1, 2003.

Carnoy, M. y P. McEwan, 1997, "Public Investments or Private Schools? A Reconstruction of Educational Improvements in Chile", Standford University.

CONEC, 2002, Situación de la Educación en el Paraguay, Consejo Nacional de Educación y Cultura, Asunción, p. 35.

CONEC, 2000, Informe sobre la Situación de la Educación Paraguaya, Consejo Nacional de Educación y Cultura, Asunción, p. 118.

Cohen, David y Deborah L. Ball, 2001, Making Change: Instruction and its improvement, Kappan, Vol 83 (1), pp. 73-77.

DGEEC, 2000, Perfil Educativo de la Población en edad de Trabajar, Dirección General de Estadística, Encuestas y Censos, Asuncion.

Fuller, Bruce and Prema Clarke, 1994, "Raising School Effects While Ignoring Culture, Local Condition and the Influence of Classroom Tools, Rules, and Pedagogy", Review of Educational Research, Vol. 64 (1).

Levin, Henry, 1999, "The Public Private Nexus in Education", American Behavioral Scientist, Vol 43 (1), p. 124-137.

Lockheed, M. and A. Verspoor, 1991, "Improving Primary Education in Developing Coun-

tries", Oxford University Press - World Bank, New York, U.S.

Nickson, Andrew, 2005, Employment Policy in an Open Economy: Paraguay, Political Economy module (First Draft), Department of Economics, University of Toronto.

OECD-WEI, 2003, Financing Education – Investments and Returns. 2002 Edition, OECD-UNESCO, Paris.

OECD, 2002, Education at a glance. OECD Indicators 2002. OECD, Paris.

PNUD, 2003, Informe Nacional sobre Desarrollo Humano. Paraguay 2003, PNUD-DGEEC, Asunción, pp. 185.

Psacharopulos, George and Patrinos, Harry, 2002, "Returns to Investments in Education: A further update, World Bank Policy Research Working Paper 2881, Washington DC., pp. 28.

Schiefelbein, Ernesto, 2005, Education and Employment in Paraguay: Issues and Perspectives, University of Toronto, CIS/CADEP, University of Toronto.

Schiefelbein, Ernesto, 2003, "Uso de la información generada por la evaluación del aprendizaje (o rendimiento) para proponer políticas educativas pertinentes", UNESCO-OREALC, Noviembre, Santiago de Chile, p. 26.

Schiefelbein, Ernesto and Jose Brunstein, 2003, Paraguay. Education and Employment, CIS/CADEP Center for Internacional Studies (CIS) de la Universidad de Toronto y el Centro de Análisis de la Economía Paraguaya (CADEP). September 2003.

Schiefelbein, Ernesto and Noel McGinn, 1980, How educational statistics can be utilized for policy analysis, in G. Lewis (ed), "Data Analysis for educational Planning", Harvard University.

SNEPE, 2002, ¿Cuanto aprenden nuestros niños y niñas? 3° Grado- 6° Grado, Informe SNEPE 2001-2002, Ministerio de Educación y Cultura, Sistema Nacional de Evaluación del Proceso Educativo, Asunción, Paraguay, p. 103.

SNEPE, 1999, Resultado de las Pruebas Nacionales, Sexto Curso, Informe SNEPE 1999, Ser-

vicio Nacional de Evaluación del Proceso Educativo, Ministerio de Educación, Asunción, p. 109. UNESCO, 2008, A View Inside Primary Schools, A World Education Indicators (WEI) cross-national study, Montreal.

UNESCO-OREALC, 2001, Situación Educativa de América Latina y el Caribe,1980-2000, Proyecto Principal de Educación, UNESCO, Santiago de Chile, p. 633.

Wolff, L.., Schiefelbein, E., y P. Schiefelbein, 2002, Primary Education in Latin America: the Unfinished Agenda, BID, Washington DC.

World Bank, 2002, World Development Indicators, The World Bank, Washington DC.

Appendix

Descriptive statistics

Table 1: Descriptive quantiles, Math scores

Quantile	1	2	3	4	5
Score mathematics	7.55	13.01	18.02	22.54	26.65
Father with up to secondary education (%)	0.54	0.54	0.57	0.6	0.58
Mother with up to secondary education (%)	0.52	0.54	0.57	0.59	0.56
Father with up to tertiary education (%)	0.05	0.06	0.05	0.05	0.07
Mother with up to tertiary education (%)	0.04	0.06	0.05	0.07	0.08
Good classroom conditions (%)	0.84	0.88	0.9	0.91	0.94
Urban area (%)	0.47	0.53	0.53	0.47	0.44
Never failed to be promoted %)	0.49	0.57	0.63	0.66	0.68
Ilness rate < 11 years, last 90 days (%)	0.56	0.56	0.55	0.54	0.56
Health insurance rate < 11 years (%)	0.19	0.2	0.21	0.21	0.21
Teacher speaks usualy Spanish (%)	0.81	0.86	0.88	0.9	0.92
Familiy only speaks Guarani (%)	0.33	0.31	0.31	0.36	0.38
Frequently works for money (%)	0.37	0.31	0.27	0.23	0.19
Frequently absent due to work (%)	0.25	0.2	0.17	0.17	0.15
Low or moderate absence rates (%)	0.52	0.61	0.66	0.68	0.72
Population in extreme poverty (% - district mean)	0.14	0.15	0.15	0.16	0.17
Population in poverty (% - district mean)	0.32	0.33	0.33	0.34	0.34
Stunting risk (% - district mean)	0.24	0.23	0.24	0.24	0.26
Stunting (% - district mean)	0.11	0.12	0.12	0.12	0.13
Starving risk (% - district mean)	0.19	0.18	0.18	0.18	0.19
Starving (% - district mean)	0.09	0.09	0.09	0.09	0.08

Source: Joint database based on EPH 2005 (DGEEC), SNEPE 2003 (MEC), WEI-SPS 2006 (UNESCO)

Descriptive statistics

Table 2: Descriptive quantiles, Communication scores

Quantile	1	2	3	4	5
Score language	6.37	10.54	13.95	17.41	22.1
Father with up to secondary education (%)	0.57	0.57	0.56	0.57	0.57
Mother with up to secondary education (%)	0.56	0.57	0.56	0.54	0.55
Father with up to tertiary education (%)	0.04	0.05	0.06	0.06	0.07
Mother with up to tertiary education (%)	0.04	0.05	0.06	0.06	0.09
Good classroom conditions	0.85	0.87	0.9	0.93	0.93
Urban area (%)	0.42	0.48	0.55	0.53	0.46
Never failed to be promoted %	0.49	0.55	0.64	0.68	0.68
Illness rate < 11 years, last 90 days (%)	0.56	0.56	0.56	0.54	0.55
Health insurance rate < 11 years (%)	0.18	0.2	0.23	0.22	0.2
Teacher usualy speaks Spanish (%)	0.8	0.86	0.88	0.9	0.93
Familiy only speaks Guarani (%)	0.39	0.35	0.28	0.32	0.36
Frequently works for money (%)	0.36	0.33	0.26	0.22	0.18
Frequently absent due to work (%)	0.25	0.21	0.17	0.15	0.16
Low or moderate absence rates (%)	0.55	0.61	0.65	0.67	0.71
Population in extreme poverty (% - district mean)	0.14	0.15	0.15	0.15	0.18
Population in poverty (% - district mean)	0.31	0.33	0.34	0.34	0.35
Stunting risk (% - district mean)	0.24	0.23	0.23	0.24	0.25
Stunting (% - district mean)	0.12	0.12	0.12	0.11	0.13
Starving risk (% - district mean)	0.19	0.18	0.18	0.18	0.19
Starving (% - district mean)	0.09	0.09	0.09	0.09	0.09

Source: Joint database based on EPH 2005 (DGEEC), SNEPE 2003 (MEC), WEI-SPS 2006 (UNESCO)

Descriptive statistics

Table 3: Table A1: Descriptive schooling statistics Paragauay

Population (1,000)	Male	Female
Preschool age, 2006	221	213
Primary school age, 2006	429	415
Secondary school age, 2006	408	395
Total population, all ages, 2007	3096	3031
Official school age (years)	**Graduation age**	**Duration**
Preschool, 2005	5	3
Primary school, 2005	11	6
Secondary school, 2005	17	6
Compulsory education, 2005	14	9
Net enrolment ratio (%)	**Male**	**Female**
Preschool NER, 2004	29.5	30.4
Primary school NER, 2004	93.6	94
Secondary school NER, 1991	25.6	27
Gross enrolment ratio (%)	**Male**	**Female**
Preschool GER, 2004	34.1	34.5
Primary school GER, 2004	113.4	109.9
Secondary school GER, 2004	66	67.3
Entrance and transition (%)	**Male**	**Female**
Primary net intake rate, 1991	72.8	71.4
Primary gross intake rate, 2004	117.2	114.3
Primary entrants with ECCE, 2004	74	76.4
Transition rate primary-secondary, 2003	90.3	90.3
Repetition and completion	**Male**	**Female**
Primary repetition rate (%), 2004	7.9	5.4
Secondary repetition rate (%), 2004	1.2	0.6
Survival rate to grade 5 (%), 2003	79.4	83.3
Survival rate to last primary grade (%), 2003	74.4	78.7
Primary completion rate (%), 2004	93.6	95.3
School life expectancy (years), 2004	10.9	10.7
Total public expenditure on education		
As % of GDP, 2003	4.7	
As % of total government expenditure, 2003	11	

Data sources: Population and Education
 United Nations Population Division, World Population Prospects: The 2006 Revision, March 2007.
UNESCO Institute for Statistics, January 2008
 http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx

Ordinary least squares regressions (baselines, individual and household characteristics)

Table 4: Baseline OLS reg. test score Mathematics, only individual and household characteristics.

Variable	Coefficient	(Std. Err.)
(Intercept)	14.4317***	0.9844
Age	-0.0489	0.0903
Female	0.0343	0.1567
Perspectives	0.3837*	0.2137
Secondary Education Father	0.5816***	0.2187
Secondary Education Mother	0.3278	0.2074
College Education Father	0.0673	0.4047
College Education Mother	1.1249**	0.3649
Parents and Homework	1.3131***	0.2494
Parents as coeducators	1.5672***	0.4189
Family language - Guaraní	0.9247***	0.2928
Media	0.9263***	0.1234
Brother / Sister look after	-0.9606***	0.2726

N		9.655
Population Size		120,670
R^2		0.079
F		9.21

Source: Own estimates based on joint database EPH 2005 (DGEEC),
SNEPE 2003 (MEC), WEI-SPS 2006 (UNESCO).

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Table 5: Baseline OLS reg. test score Communication, only individual and household characteristics.

Variable	Coefficient	(Std. Err.)
(Intercept)	12.7322***	0.8113
Age	-0.1393	0.0765
Female	0.2895**	0.127
Perspectives	0.6413	0.1605
Secondary Education Father	0.3426**	0.1587
Secondary Education Mother	0.0515	0.1648
College Education Father	0.6789**	0.2834
College Education Mother	0.4409	0.2886
Parents and Homework	1.1158***	0.1958
Parents as coeducators	0.8424**	0.3414
Family language - Guaraní	0.3003	0.2555
Media	0.7751***	0.0921
Brother / Sister look after	-0.7655***	0.216

N		9.655
Population Size		116,145
R^2		0.065
F		10.63

Source: Own estimates based on joint database EPH 2005 (DGEEC),
SNEPE 2003 (MEC), WEI-SPS 2006 (UNESCO).

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Ordinary least squares regressions, full specification

Table 6: Complete OLS regression test score Mathematics

Variable	Coefficient	(Std. Err.)
(Intercept)	15.6073***	1.6811
Age	-0.0001	0.0889
Female	-0.145	0.1574
Perspectives	0.3981*	0.2134
Work	-0.799***	0.2552
Work Money	-1.3013***	0.2182
Secondary Education Father	0.4983**	0.214
Secondary Education Mother	0.1458	0.2091
College Education Father	0.0574	0.3875
College Education Mother	1.0822***	0.3532
Parents and Homework	0.9426***	0.2429
Parents as coeducators	1.4883***	0.4055
Family language - Guaraní	0.7621***	0.2914
Media	0.738***	0.1104
Brother / Sister look after	-0.8708***	0.268
Urban area	-0.2817	0.503
Public sector	-0.591	0.5572
District health insurance rate	2.1826*	1.1345
District sickness rate	-4.0249**	1.5734
District malnutrition rate	-5.5158**	2.6295
School Language - Castellano	1.2945***	0.4131
Optimal classroom conditions	1.4993***	0.3928
Small school	1.1765***	0.5038
Alto Paraná	-2.5122***	0.7673
Caazapá	2.62***	1.062
Canindeyú	3.8905***	1.3961
Concepción	-1.8352*	0.998
Cordillera	-3.4031***	0.7658
Misiones	-1.9002*	1.0966
Paraguarí	1.2318	0.8359

N	9.444
Population Size	119,850
R^2	0.1188
F	9.31

Source: Own estimates based on joint database EPH 2005 (DGEEC),
SNEPE 2003 (MEC), WEI-SPS 2006 (UNESCO).

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Table 7: Complete OLS regression test score Communication

Variable	Coefficient	(Std. Err.)
(Intercept)	14.3255***	1.3302
Age	-0.0837	0.0723
Female	0.1634	0.1279
Perspectives	0.633***	0.1607
Work	-0.4069**	0.1882
Work Money	-1.1442***	0.1672
Secondary Education Father	0.327**	0.1554
Secondary Education Mother	-0.042	0.1659
College Education Father	0.6636**	0.2779
College Education Mother	0.3311	0.2705
Parents and Homework	0.8589***	0.1922
Parents as coeducators	0.788**	0.3424
Family language - Guaraní	0.2751	0.2522
Media	0.6185***	0.0864
Brother / Sister look after	-0.6818***	0.2127
Urban area	-0.0072	0.4222
Public sector	-0.8196*	0.4412
District health insurance rate	0.554	0.8475
District sickness rate	-3.8469***	1.3201
District malnutrition rate	-2.8985	2.107
School Language - Castellano	1.0687***	0.3399
Optimal classroom conditions	0.9436***	0.3113
Small school	0.4379	0.4323
Alto Paraná	-1.8693***	0.5929
Caazapá	1.2282	1.0663
Canindeyú	2.2325*	1.2831
Concepción	-1.7225**	0.7974
Cordillera	-1.7909***	0.6714
Misiones	-1.1124	0.7932
Paraguarí	0.703	0.8617

N	9.444
Population Size	115,286
R^2	0.098
F	9.69

Source: Own estimates based on joint database EPH 2005 (DGEEC),
SNEPE 2003 (MEC), WEI-SPS 2006 (UNESCO).

Significance levels : $*: 10\%$ $**: 5\%$ $***: 1\%$

Quantile regressions (Koenker and Basset, 1978)

Table 8: Quantile Regression, dependent variable: Math scores

| Variable	Q(0.1)	Q(0.25)	Median	Q(0.75)	Q(0.9)
(Intercept)	8.03546***	10.29893***	15.38378***	20.30685***	24.92299***
Age	-0.11795	-0.06398	-0.02463	0.0658	0.0348
Female	0.08544	0.01587	-0.32642	-0.29627	-0.35571**
Motivation	0.30428	0.04793	0.58662**	0.42608**	0.10436
Work	-0.82252***	-1.01847***	-1.06015***	-0.6656***	-0.38651*
Work Money	-1.1497***	-1.74203***	-1.7***	-1.09333***	-0.94389***
Secondary Education Father	0.84632***	0.9744***	0.644**	0.24883	0.10963
Secondary Education Mother	0.04949	0.28055	0.18983	-0.1551	-0.02582
College Education Father	0.07386	0.40913	-0.12261	0.06102	0.01331
College Education Mother	0.60621	0.92914	1.33207**	0.80413**	0.65121**
Parents and Homework	0.38122	1.09165***	1.46262***	0.79098***	0.35884**
Parents as coeducators	1.0703**	1.86552***	1.97967***	1.02659**	0.69359**
Family language - Guaraní	0.21789	0.62825**	1.16788***	0.85898***	0.3184*
Media	0.44829***	0.80482***	1.10758***	0.76661***	0.36665***
Brother / Sister look after	-0.82258***	-1.03402***	-1.40344***	-0.76387***	-0.28006*
Urban area	0.04433	-0.09904	-0.46376*	-0.36185	-0.53543***
Public sector	-0.60093*	-0.70235**	-0.38606	-0.34692	-0.49659**
District health insurance rate	2.44264***	2.20781***	1.59737**	1.66835**	1.33493***
District sickness rate	-2.97726***	-3.24105***	-5.08069***	-3.23731***	-2.14138***
District malnutrition rate	-3.9**	-7.43097***	-5.16943***	-5.3514***	-3.13623***
School Language - Castellano	1.03103***	1.33521***	1.68139***	1.18315***	0.932***
Optimal classroom conditions	1.47226***	1.65958***	1.50929***	1.81059***	1.21399***
Small school	0.49039	1.28924***	1.88915***	0.99451***	0.42288**
Alto Paraná	-2.37316***	-3.44164***	-2.95264***	-1.80752***	-1.4043***
Caazapá	2.77513***	3.74649***	2.08667***	2.25112***	1.6299***
Canindeyú	5.20881***	6.65209***	3.12545***	3.37199***	2.30934***
Concepción	-1.31932**	-1.94716***	-2.9028***	-1.21969***	-0.83973***
Cordillera	-1.84356***	-3.36033***	-4.0396***	-3.85493***	-2.80983***
Misiones	-0.85852	-1.65761**	-2.53836***	-1.41223**	-1.23454***
Paraguarí	0.74857	1.20832*	1.6507***	1.28937***	0.95563***
Observations	9444	9444	9444	9444	9444
Pseudo R-squared	0.0501	0.0758	0.0803	0.0562	0.0494
Bootstrapped Std. Errors		500 Repetitions			

Source: Own estimates based on joint database EPH 2005 (DGEEC),

SNEPE 2003 (MEC), WEI-SPS 2006 (UNESCO).

Cluster-robust standard errors.

Significance levels : * : 10%   ** : 5%   * * * : 1%

Table 9: Quantile Regression, dependent variable: Communication scores

Variable	Q(0.1)	Q(0.25)	Median	Q(0.75)	Q(0.9)
(Intercept)	8.51413***	11.46177***	14.4894***	17.18069***	20.96956***
Age	-0.16179**	-0.22362***	-0.15387*	0.04318	0.04651
Female	-0.13539	0.0511	0.34497**	0.26096	-0.12669
Perspectives	0.36665**	0.47409***	0.65273***	0.66997***	0.56851***
Work	-0.18208	-0.49329***	-0.63483***	-0.5847**	-0.16699
Work Money	-0.72753**	-1.09133***	-1.29909***	-1.21723***	-1.37284***
Secondary Education Father	0.17423	0.31992*	0.43211**	0.20597	0.11052
Secondary Education Mother	-0.05242	-0.03494	-0.0441	-0.038	0.11865
College Education Father	1.39227***	0.66221**	0.5963	0.28684	0.82713*
College Education Mother	-0.1171	0.31959	0.17408	0.97708*	0.53782
Parents and Homework	0.93685***	0.82595***	1.2292***	0.85354***	0.72083***
Parents as coeducators	0.63497**	0.91474***	0.93955**	0.65828*	0.56181
Family language - Guaraní	-0.22217	-0.22806	0.18257	0.69842***	0.90868***
Media	0.42372***	0.56705***	0.75762***	0.63904***	0.60414***
Brother / Sister look after	-0.60224***	-0.72816***	-0.78698***	-0.52098***	-0.39774*
Urban area	0.26103	0.30644**	0.40803**	-0.33521	-0.6416**
Public sector	-1.08585***	-1.09374***	-0.76216***	-0.83005***	-0.37136
District health insurance rate	1.28362**	1.29047***	0.4834	-0.89179	0.38528
District sickness rate	-2.49273***	-3.17164***	-4.65138***	-4.32255***	-4.85762***
District malnutrition rate	-2.33022**	-3.20336***	-3.60702***	-1.91072	-3.5774***
School Language - Castellano	0.90004***	1.14141***	0.89637***	1.42537***	1.09077***
Optimal classroom conditions	0.80664***	0.91849***	1.12471***	0.97401***	0.91666***
Small school	0.29891	0.33063*	1.10705***	0.57353**	-0.45568*
Alto Paraná	-1.45775***	-1.5737***	-2.1536***	-2.13901***	-1.84683***
Caazapá	1.54704***	1.27841***	0.60847	1.6721***	1.66985***
Canindeyú	2.39195**	3.18105***	1.75381***	1.62072**	2.11931**
Concepción	-0.47656	-1.25528***	-2.36205***	-2.36516***	-1.83***
Cordillera	-0.54142*	-1.0401***	-1.75007***	-2.80425***	-2.60528***
Misiones	-0.20753	-0.73771	-1.53501***	-1.05133*	-1.19874*
Paraguarí	0.18046	0.30585	1.00544**	1.07553**	0.79978**
Observations	9444	9444	9444	9444	9444
Pseudo R-squared	0.0406	0.0565	0.0661	0.0558	0.0503
Bootstrapped Std. Errors		500 Repetitions			

Source: Own estimates based on joint database EPH 2005 (DGEEC),

SNEPE 2003 (MEC), WEI-SPS 2006 (UNESCO).

Cluster-robust standard errors.

Significance levels :   ∗ : 10%   ∗∗ : 5%   ∗ ∗ ∗ : 1%

Graphs of quantile regression coefficients (math scores)

Figure 1: Guarani Language on Math Score

Guarani Speaker Family (Math)
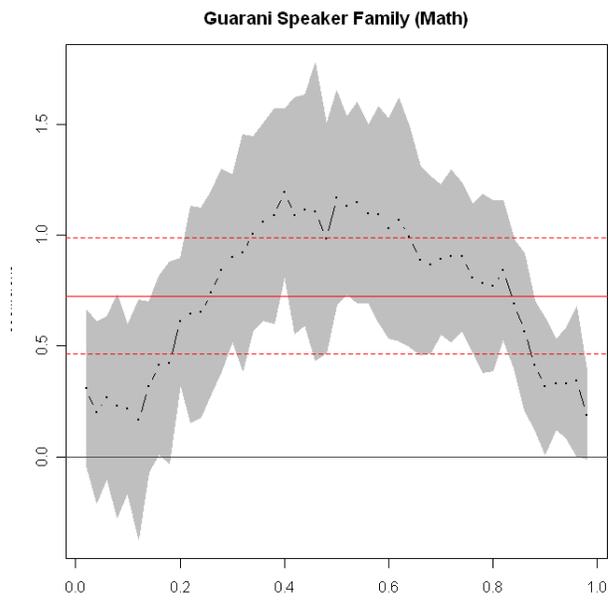
Source: Own elaboration using quantreg package in R
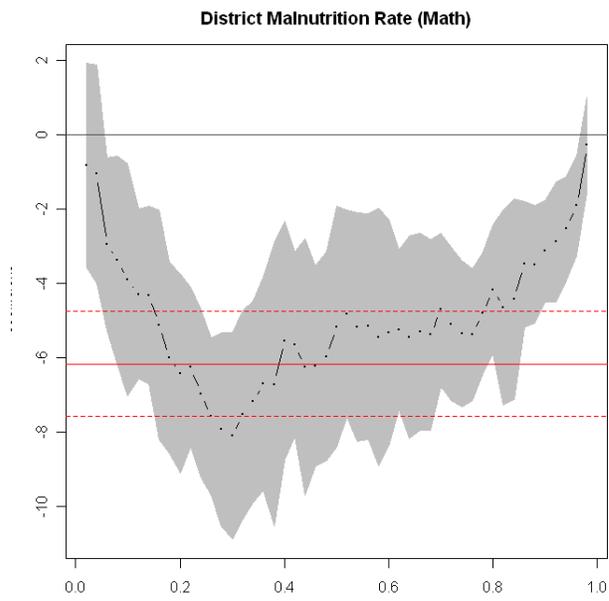
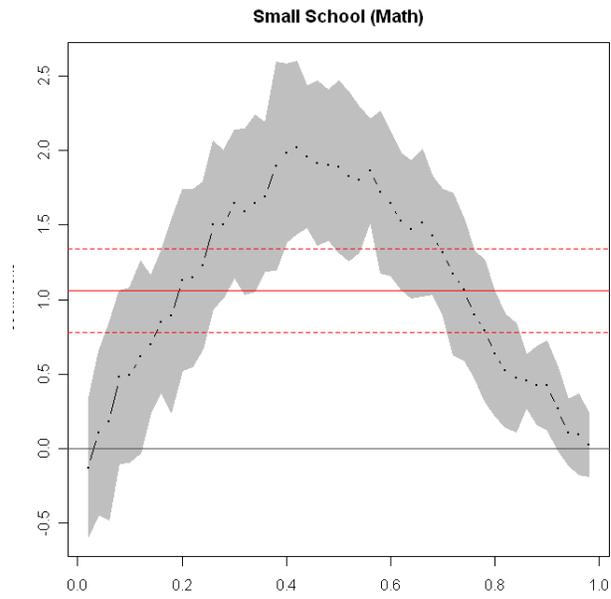Figure 2: District Health Insurance on Math Score

District Health Insurance Rate (Math)
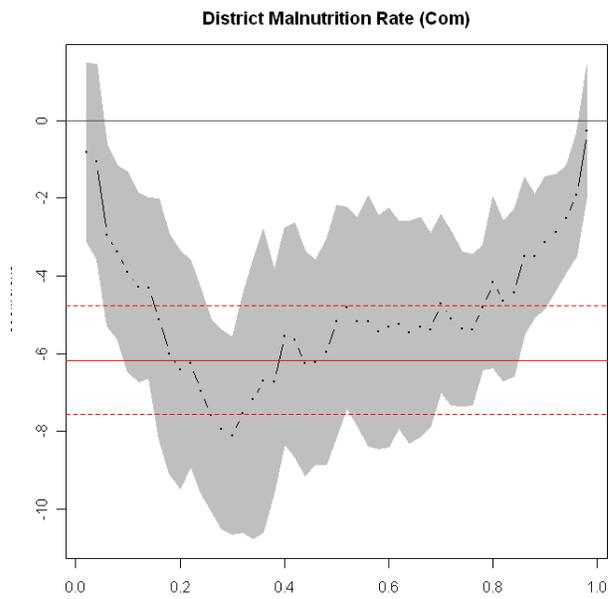

Source: Own elaboration using quantreg package in R

Graphs of quantile regression coefficients (math scores)

Figure 3: Malnutrition on Math Score

District Malnutrition Rate (Math)

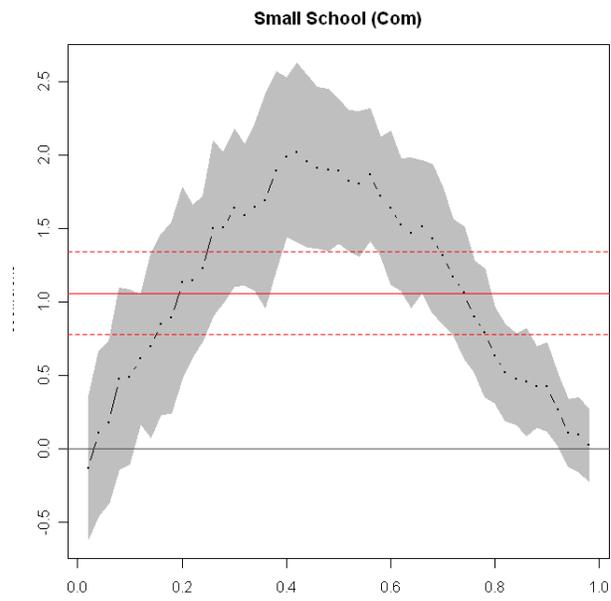Source: Own elaboration using quantreg package in R

Figure 4: Class conditions on Math Score
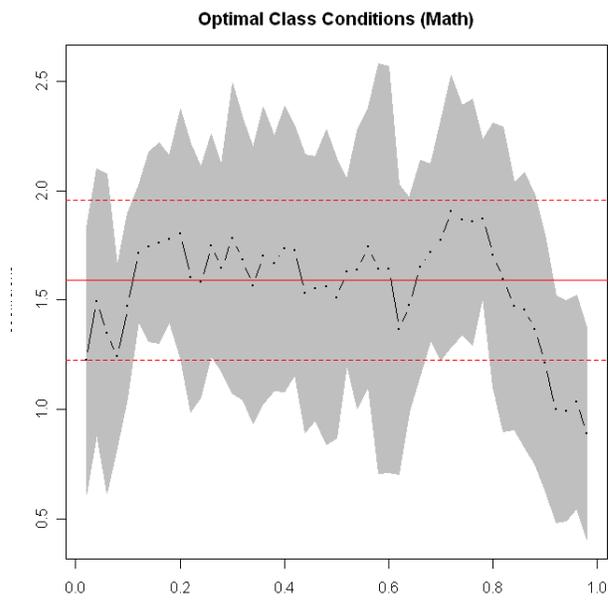
Optimal Class Conditions (Math)

Source: Own elaboration using quantreg package in R

Graphs of quantile regression coefficients (math scores)

Figure 5: Size of School on Math Score

Small School (Math)

Figure 6: Teacher's language on Math Score
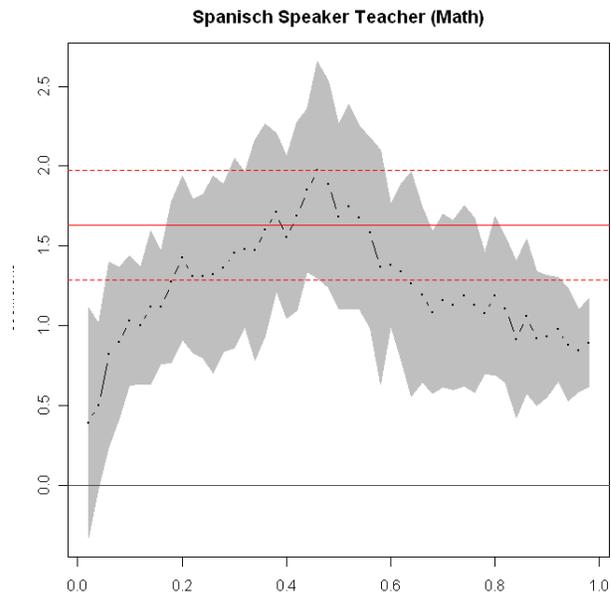
Spanisch Speaker Teacher (Math)

Graphs of quantile regression coefficients (communication scores)

Figure 7: Guarani Language on Communication Score

Guarani Speaker Family (Com)
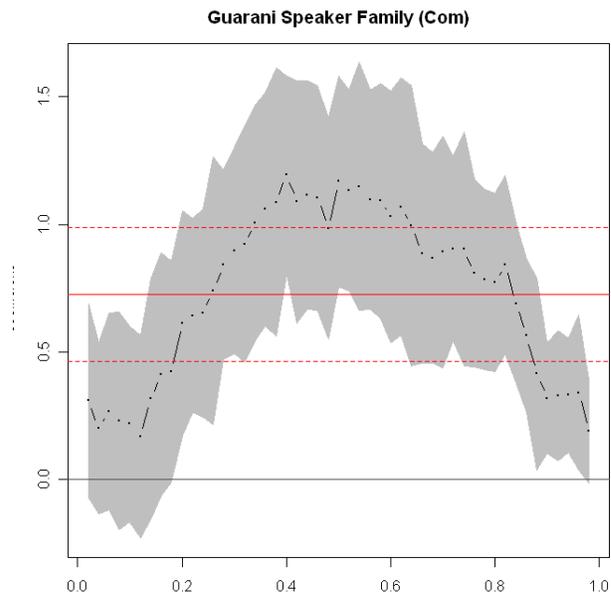
Source: Own elaboration using quantreg package in R

Figure 8: District Health Insurance on Communication Score
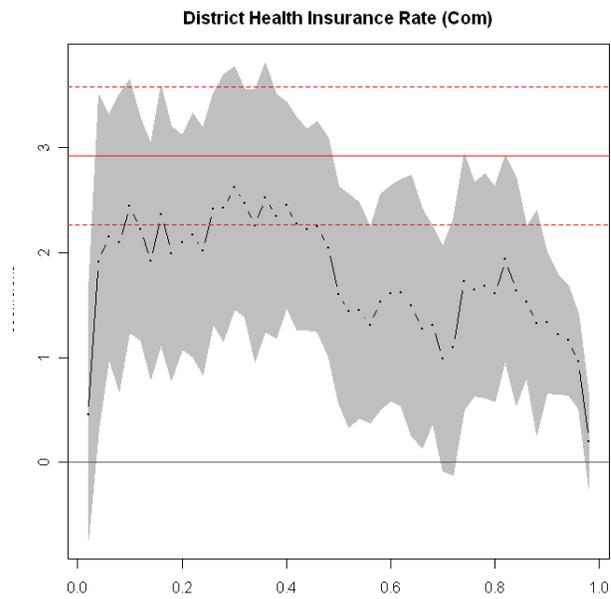
District Health Insurance Rate (Com)

Source: Own elaboration using quantreg package in R

Graphs of quantile regression coefficients (communication scores)

Figure 9: Malnutrition on Communication Score

District Malnutrition Rate (Com)

Source: Own elaboration using quantreg package in R

Figure 10: Class conditions on Communication Score
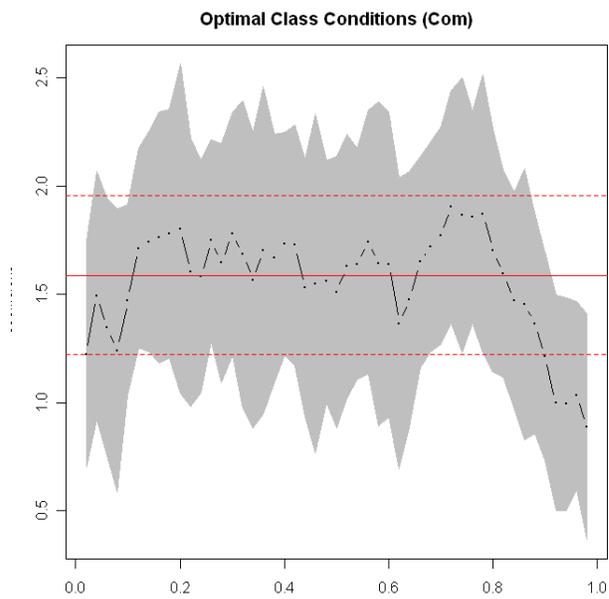
Optimal Class Conditions (Com)

Source: Own elaboration using quantreg package in R

Graphs of quantile regression coefficients (communication scores)

Figure 11: Size of School on Communication Score

Small School (Com)

Source: Own elaboration using quantreg package in R

Figure 12: Teacher's language on Communication Score

Spanisch Speaker Teacher (Math)
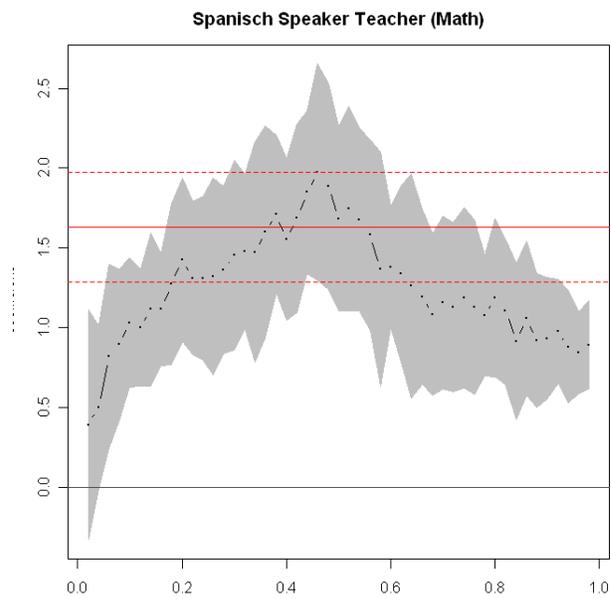
Source: Own elaboration using quantreg package in R

Descriptive Graphs

Figure 13: Paraguayan pre primary enrollment rates
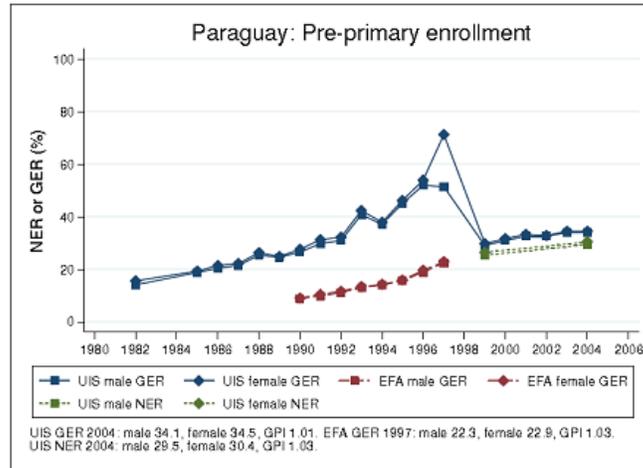
Source: UNICEF, Division of Policy and Practice, Statistics and Monitoring Section,
www.childinfo.org, May 2008

Figure 14: Paraguayan Primary attendance/enrollment rates
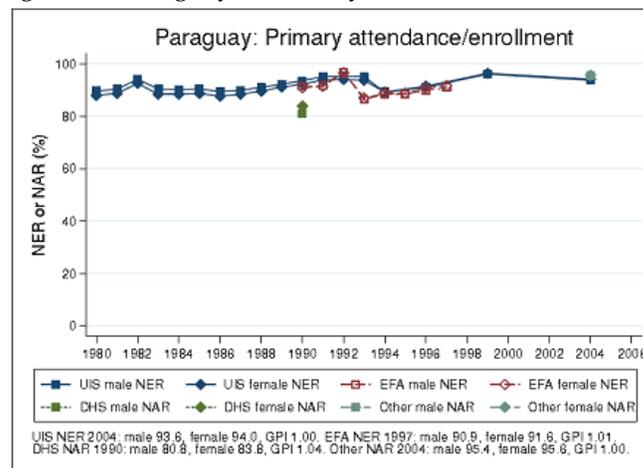
Source: UNICEF, Division of Policy and Practice, Statistics and Monitoring Section,
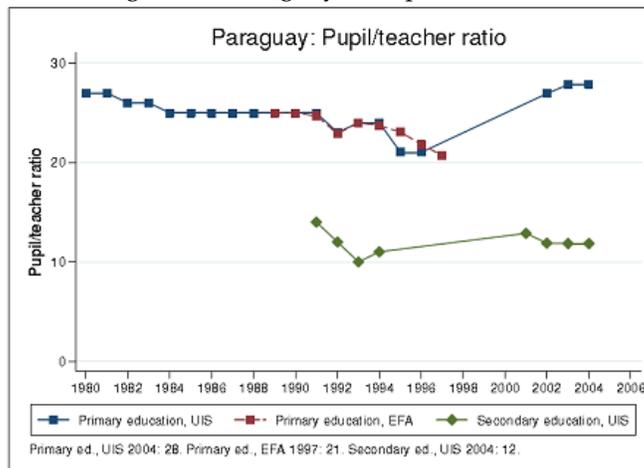www.childinfo.org, May 2008

Descriptive Graphs

Figure 15: Paraguayan Secondary attendance/enrollment rates
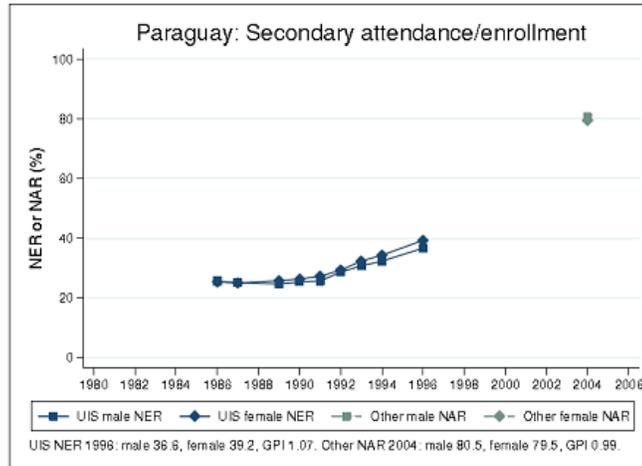
Source: UNICEF, Division of Policy and Practice, Statistics and Monitoring Section, www.childinfo.org, May 2008
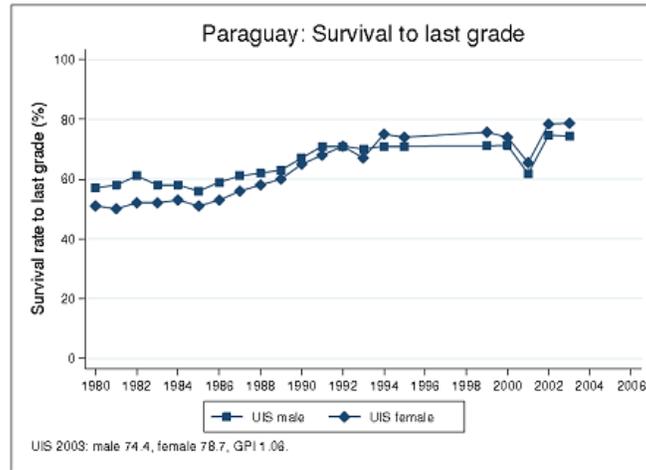
Figure 16: Paraguayan Pupil/teacher ratio

Source: UNICEF, Division of Policy and Practice, Statistics and Monitoring Section, www.childinfo.org, May 2008
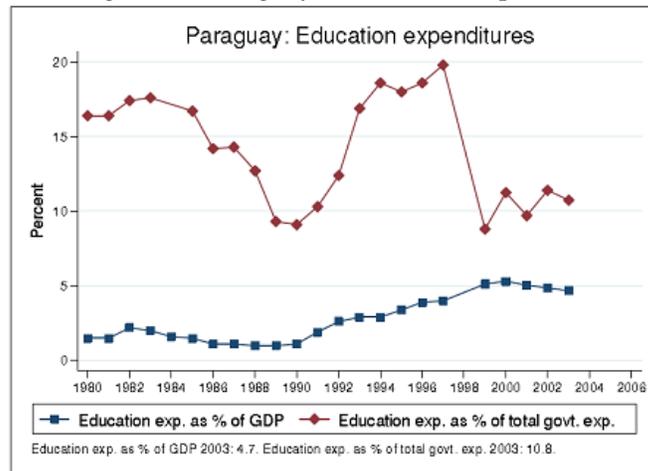
Descriptive Graphs

Figure 17: Paraguayan Survival to last grade rates

Source: UNICEF, Division of Policy and Practice, Statistics and Monitoring Section,
www.childinfo.org, May 2008

Figure 18: Paraguayan education expenditures

Source: UNICEF, Division of Policy and Practice, Statistics and Monitoring Section,
www.childinfo.org, May 2008