Educational quality and regional disparity: one application of hierarchical model to Brazilian dataset

Tatiane Menezes
(Dep. Economia/PIMES/UFPE e Pesquisadora CNPq)
Sammara Cavalcanti Soares
(Dep. Economia/PIMES/UFPE)

Abstract

The quality of the educational system is crucial to the most diverse socioeconomic and cultural aspects of a country. The objective of this paper is highlighting the importance of quality of education to explain the wide income disparities between Brazilian people. In face of such evidence, and considering the need to better understand the factors associated with education quality and efficiency, this study aims to investigate which factors are the most significant to determine elementary school students' performance in capital and small towns. For this purpose, it will be used the Brazilian Educational Search (SAEB) for 4th grades of elementary education. The estimation method is the hierarchical regression, which, in this work, will be structured into 2 levels, respectively: students, and school. The results suggest, after socioeconomic control, the positive correlation of principal's characteristic; class-time, kindergarten and presence of library to a better performance. In opposite way, the influence of professor's turnover seems to be quite harmful to determine the student´s score in capital city and small towns respectively. These results can serve as aid to public policies aimed at equity and educational qualification.

Key words: hierarchical model, educational accountability; spatial inequality
JEL: A2, H52, H75 I21, C4, R12

INTRODUCTION

The understanding of the role of schools on their pupils` performances is a line of research which has been given increased importance by the international literature ((Raudenbush e Bryk, 1986; Willms J. D. 1984; Sammons, P., Hillman, J., Mortimore, P.,1995). Its importance comes, among other aspects, from the relationship observed between economic development and population educational quality. In developing countries, this concern has been increasingly high. Evaluating schools` participation on pupils` performance, it is possible to create effective public policies to improve the country's human capital and, therefore, its capacity for growth.

Papers such as Barros and Reis (1990), suggest that part of income inequality between countries and between regions within the same Country is due to the difference observed between their respective educational frameworks. Areas with better schools would form a more qualified human capital and thus higher input for its development. In this context, the concern of this study consists of identifying what is the school's role on pupils` performance in rich and poor regions.

The difficulty of this work lies in the measurement of the impact of the quality of school on learning performance (Hanushek, 1971; Richard Murnane, 1975; David Armor et al., 1976; Albert Park and Emily Hannusm, 2002; Claudia Uribe et al., 2003). In an attempt to isolate the school effect, literature usually employs database composed with characteristics of both the students and schools. This composite database allows to separate the fixed effect of students on examinations grades, so that the resultant impact would be assigned to the school.

The numerous attempts to understand the relationship between school`s quality and pupils' learning show different conclusions and methodologies. The only result common between them all is the huge importance of socioeconomic characteristics of households. But, even if such family factors play a strong influence on students' learning, articles such as Menezes Child (2007), Willms and Somers (1999) and Rivkin, Hanushek and Kain (2005) show that characteristics from schools, teachers and even from classes are responsible for a substantial contribution in the student's proficiency.

A possible definition of effective school, perhaps the most simple, is presented by Mortimore (1991). According to this author, effective school is the one where the student's progress goes beyond what would be expected if taken into account his/her socioeconomic characteristics at the time he/she joins school. Indeed, estimates of the effect-school showed the important role that it can have not only on pupils` cognitive development but also on its social function accomplishment while promoting population social leverage by removing such children and youth of the vicious cycle of poverty.

One hypothesis often raised by the literature, as well as usually accepted in the theory of production, is the concavity of school` s production function in relation to its inputs. In this case, students of poorer resources schools could benefit significantly with the increase of more and better school inputs. (Felício and Fernandes, 2005). Supporting this hypothesis, Hanushek et al. (1996), after analyzing the rural area of the Northeast of Brazil (characterized by great economic difficulty), suggested that the infrastructure of educational units; as well as the presence of notebooks, audiovisual resources and materials; are positively related to student performance in this region.

Seeing the multitude of factors that operates in the determination of school's performance, as well as its complex interactivity among the different levels of influence, this study seeks to throw light on the quality of education through hierarchical models of regression. The variability of grades among students from 4th grade (Elementary Education) will be investigated trying to observe where schools play a greater influence on learning proficiency, if in the richest (capital) or the poorest (not capital) regions.

Sammons et al (1995) lists 11 determinant characteristics of effective schools. Among them we can mention the appropriate learning environment, strong principal`s leadership, high teacher's expectation regarding student performance, frequent student monitoring in front of his/her learning progress within the classroom and home-school. There is robust empirical evidence between better academic performance and the average time of stay of children and youngsters in schools; as well as frequenting pre-school and doing the homework (see Curi and Menezes, 2006).

The main difficulty with this kind of study is the quality of the database, mainly in developing countries where only a small amount of the resources are designed to mount the database. The National Basic Education Evaluation System (SAEB: translation from portuguese), database for this abstract was developed by the Ministry of Education`s National Study & Educational Research Institute (INEP: translation from portuguese). This system evaluates the 4th and 8th grades (Elementary Education) and the 3rd year (High School) students` performances on Portuguese and Mathematics disciplines from public and private schools of all the Brazil`s states. Through questionnaires, applied in parallel to the exams, it is possible to collect information about students` social, economic and cultural contexts as well as teachers', principals' and schools physical structures` characteristics. The sampling procedures are based on scientific methodology that assures precision on population parameters estimates.

The state of Pernambuco`s 4th grade pupils` Mathematics grades database obtained from SAEB (2007) is used on this paper. This state's educational and socioeconomic context favored it be chosen as target of research. First, Pernambuco reflects the context of the Brazil`s Northeast region, which present, historically, economic and social indexes rather unfavorable, compared to the South and Southeast regions. The last ones, in turn, concentrated much of national studies on the quality of education, therefore, not by chance, the Northeast region, and particularly Pernambuco, remain in deficit of more in-depth studies scholars in the educational quality area, even being an essential issue for the economic and social development of the region. Second, but not least, is the fact that Pernambuco submitted, in recent years, one of the worst educational indices of Brazil, with the worst IDEB[1], in 2006, for final elementary schools` years, in addition to high rates of flunking and dropout (an average of 36% flunking and abandonment in elementary school) and high percentages of age-grade distortion (above 50% in elementary school and 70% in high school). Furthermore, Pernambuco holds the largest GDP of Northeast[2] and, however, presents school results as bad or worse than the other states in this region, especially regarding the 8th grade and 3rd year of high school.

The point of view of this article, when comparing the impact of schools in the capital front those who are not in the capital, add to the studies of the area an alternative way to investigate the influence of school factors on students resourcefulness, in addition to enrich the range of estimates that already exist in the literature about the topic. Furthermore, it aims to realize if there are great similarities and/or differences between the estimates of capital and non-capital, since the capital of the state (Recife) is historically richer and more productive.

In addition to this introduction, this article is divided into 5 sections. In the second section, there is a descriptive analysis of data; in the third, the methodological procedures that were applied are discussed; in the fourth, results found are highlighted and discussed and in the fifth and last part holds the studies conclusion.

3. EMPIRICAL STRATEGY

The school systems are a typical example of the hierarchical structure, since the students are nested in classes, the classes grouped in schools, schools grouped in a determined site and there forth (SOARES, T. e MENDONÇA, M. 2003). Thus, the role of factors from different levels of influence on student performance indicates the necessity of using multilevel regression models, which have become standard in empirical educational research. (GOLDSTEIN, 2001, p. 86 apud ANDRADE and SOARES2008).

In this study, are considered only models with two levels: students and schools. students

[1] Basic education development índex (IDEB – translated from portuguese)

[2] Source: GDP for 2007

(level 1) are represented by index i, and schools (level 2) are identified by the index j. The variable response (y_{ij}) of the model is the proficiency of student i who belongs to the school j, and the explanatory variables (x_{2ij} until x_{pij}) associated to the student, serve as a socioeconomic control for their analysis of the model (see table). The classic regression model specifies the following relation:

$$y_{ij} = \beta_1 + \beta_2 x_{2ij} \ldots + \beta_p x_{pij} + \xi_{ij} \qquad (1)$$

where x_{2ij} through x_{pij} are covariates and ξ is a residual.

It may be unrealistic to assume that the student math score of children at the same school are independent given observed covariates, or in other words, that the residuals, ξ_{ij} and $\xi_{i'j}$ are independent. We can therefore split the total residual or error into two error components:

$$\xi_{ij} = \zeta_j + \varepsilon_{ij}$$

Substituting ξ into the multiple regression models (1) we obtain a linear random intercept model with covariates.

$$\begin{aligned} y_{ij} &= \beta_1 + \beta_2 x_{2ij} \ldots + \beta_p x_{pij} + \zeta_j + \varepsilon_{ij} \\ &= \beta_1 + \zeta_j + \beta_2 x_{2ij} \ldots + \beta_p x_{pij} + \varepsilon_{ij} \end{aligned} \qquad (2)$$

This model can be viewed as a regression model with a school specific intercept $\beta_1 + \zeta_j$. The random intercept ζ_j can be considered a random parameter that is not estimate along with fixed parameters β_1 through β_p, but the ζ_j variance (ψ) is estimated together the variance θ of the ε_{ij}. The linear random intercept model with covariates is the simplest example of a linear mixed model where there are both fixed and random effects.

The random intercepts or level-2 residual ζ_j is a school specific error component, which remains constant across student, whereas the level-1 residual ε_{ij} is a child-specific error component, which varies between students i as well as schools j. The ζ_j are independent over schools, the ε_{ij} are independent over schools and students, and two error components are independent of each other.

The school specific error component ζ_j represents the combined effects of omitted school characteristics or unobserved heterogeneity. If ζ_j is positive, the total residuals for schools j, ξ_{ij}, will tend to be positive, leading to students score than predicted by the covariates, and if ζ_j is negative the total residual will tend to be negative. Since ζ_j is shared by all responses for the same school, it induces within school dependence among the total residuals ξ_{ij}.

Letting $x_{ij} = (x_{2ij} \ldots, x_{pij})'$ be vector consisting of all observed covariates, the exogeneity assumption is:

$$E(\zeta_j \mid x_{ij}) = 0 \tag{3}$$

and

$$E(\varepsilon_{ij} \mid x_{ij}, \zeta_j) = 0 \tag{4}$$

From which it follows that $E(\varepsilon_{ij} \mid x_{ij}) = 0$. These assumption ensure that the population averaged or marginal regression is linear

$$\begin{aligned} E(y_{ij} \mid x_{ij}) &= E(\beta_1 + \beta_2 x_{2ij} \ldots + \beta_p x_{pij}) + \underbrace{E(\zeta_j \mid x_{ij})}_{0} + \underbrace{E(\varepsilon_{ij} \mid x_{ij})}_{0} \\ &= \beta_1 + \beta_2 x_{2ij} \ldots + \beta_p x_{pij} \end{aligned} \tag{5}$$

and that the cluster specific or conditional regression is linear:

Regarding distributional assumption, we specified that

$$\zeta_j \mid x_{ij} \sim N(0, \psi)$$

and

$$\varepsilon_{ij} \mid x_{ij}, \zeta_j \sim N(0, \theta) \tag{6}$$

From which it follows that: $\zeta_j \sim N(0, \psi)$ and $\varepsilon_{ij} \sim N(0,\theta)$.

If the total residual error terms are homoscedastic then the responses y_{ij}, given observed covariates $\mathbf{x}_{ij}$, are also homoscedastic.

$$\text{Var}(y_{ij} \mid x_{ij}) = \text{Var}(\xi_{ij}) = \text{Var}(\zeta_j + \varepsilon_{ij}) = \psi + \theta$$

The correlation between the total residual for any two children i and i' in the same school j, also called the residual correlation, is

$$\rho \equiv Cor(\xi_{ij}, \xi_{i'j}) = \frac{\psi}{\psi + \theta}$$

Thus rho is also the intraclass (between effect or between school effect) correlation of response y_{ij} and $y_{i'j}$ for school j, given the covariates

$$\rho \equiv Cor(y_{ij}, y_{i'j} \mid x_{ij}, x_{i'j}) = \frac{\psi}{\psi + \theta} \qquad (7)$$

It follows from the exogeneity assumptions stated in (3) and (4) that both ζ_j (school) and ε_{ij} are uncorrelated with covariates.

However, endogeneity is often discussed in terms of correlation between the error terms and covariates. For example, if kindergarten is assumed to be correlated with the random intercept for schools, this represents the effect of omitted school specific covariates on student score. We can easily relax the assumption that the between and within effects are the same for particular explanatory variables, say, x_{2ij}, by using the model:

$$y_{ij} = \beta_1 + \beta_2^1(x_{2ij} - \bar{x}_{2.j}) + \beta_2^2\bar{x}_{2.j} \ldots + \beta_p x_{pij} + \zeta_j + \varepsilon_{ij} \qquad (8)$$

Which collapse to the original random intercept model (2) if $\beta_2^1 = \beta_2^2 = \beta_2$.

The deviation from the cluster mean of $\bar{x}_{2ij}$ (for example: kindergarten mean by school) is an instrumental variable for x_{2ij} because it is correlated with x_{2ij}, but uncorrelated with random intercept ζ_j. We can also view this model as relaxing the assumption that the random intercept is uncorrelated with x_{2ij} if we think of $\left(\beta_2^2\bar{x}_{2.j} + \zeta_j\right)$ as random intercept.

It is important to note that we do not need to subtract the cluster mean $\bar{x}_{2ij}$ from x_{2ij} as long as we include the cluster mean in the model since:

$$\begin{aligned} y_{ij} &= \beta_1 + \beta_2^1(x_{2ij} - \bar{x}_{2.j}) + \beta_2^2\bar{x}_{2.j} \ldots + \beta_p x_{pij} + \zeta_j + \varepsilon_{ij} \\ &= \beta_1 + \beta_2^1 x_{2ij} + (\beta_2^1 - \beta_2^2)\bar{x}_{2.j} \ldots + \beta_p x_{pij} + \zeta_j + \varepsilon_{ij} \end{aligned} \qquad (9)$$

If x_{2ij} is cluster-mean centered affects as in the first line of (9), the coefficient of cluster mean represents the between effect, and if not as in the second line of (9) it represents the differences in between and within effects.

2. COMPARATIVE ANALISYS AND DATA DESCRIPTION

2.1 Comparative analysis

Figure 1 and 2 show the abandon and fail rates of the elementary education of Brazil, Northeast, Pernambuco and São Paulo (state of the Brazilian Southeast; model for its economic and education efficiencies) from 1999 to 2007; and figures 3 and 4 show the results provided from the age conclusion distortion from 1998 to 2005.

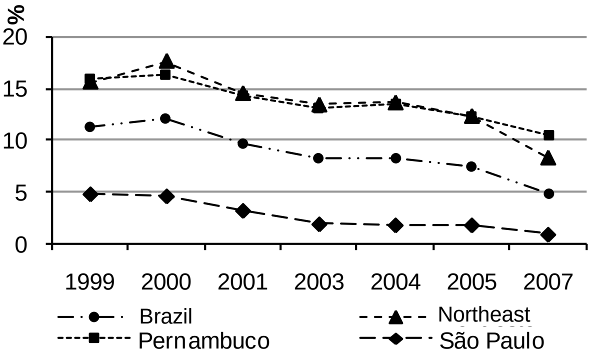


Figure 1 – Abandon Rate
Elementary Education (1999-2007)
Font: Self Elaborated. SCHOOL CENSUS data.

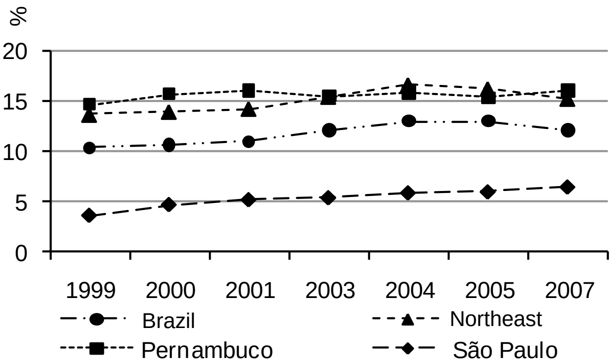


Figure 2 – Flunk Rate
Elementary Education (1999-2007)

It’s clearly noted that Pernambuco state show the highest incidences in almost every analyzed year, on both fail and abandon rates (this last, although decreasing, still reached 10% of the children in 2007).

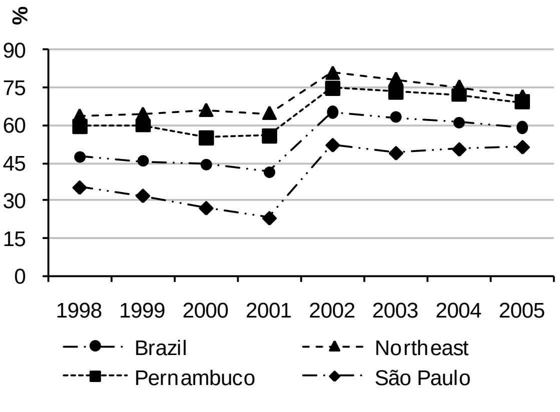


Figure 3 – Age Conclusion Distortion
Elementary Education (1998–2005)

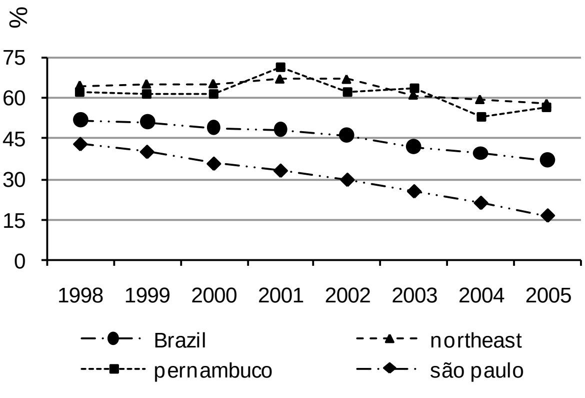


Figura 4 – Age Conclusion Distortion
High School Education (1998–2005)

Font: Self Elaborated. SCHOOL CENSUS data

Consequently, the age grade distortion reflect the anterior indexes and reach almost 75% of Pernambuco's youth at the final years of elementary education in 2005.

The following figures (5 and 6) provide information about the grades attained in Portuguese and Mathematics on all SAEB's performed in Brazil and Figure 7 reveals, more specifically, the percentage of 4th grade pupils who attained appropriate minimum performance score[3] on 2007's SAEB.

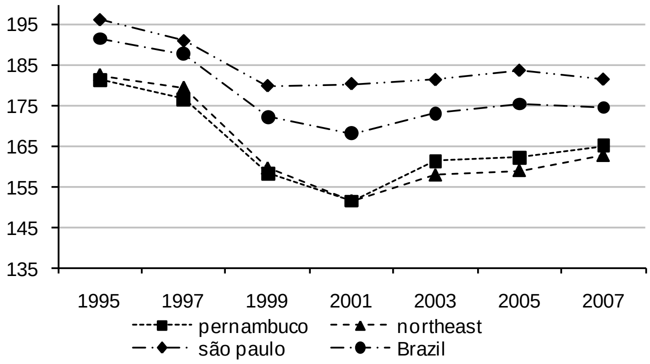


Figure 5
Portuguese performance – 4th grade (E.Educ.)
(1995) – (2007)

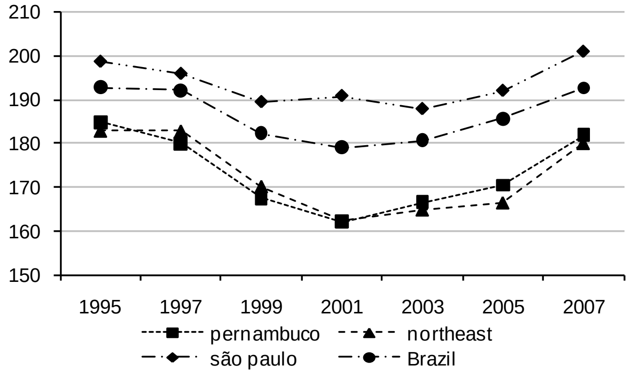
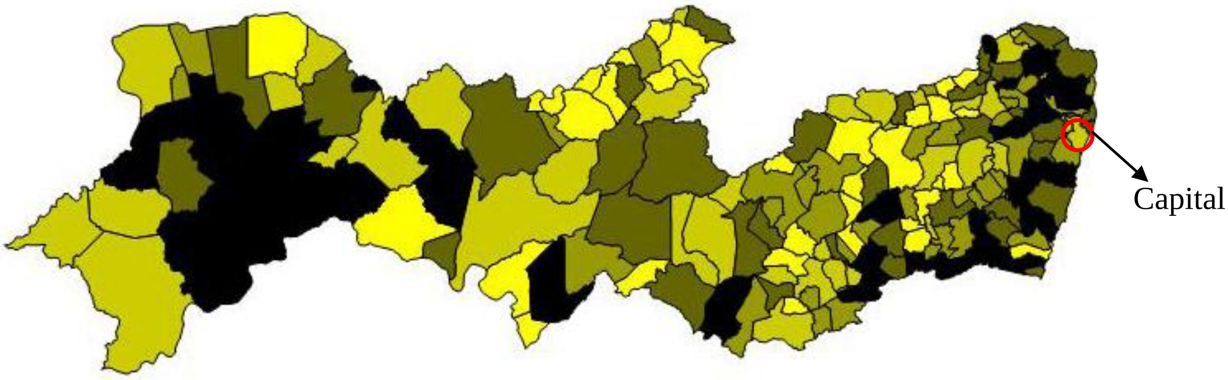


Figure 6
Mathematics performance – 4th grade (E.Educ.)
(1995 – 2007)

Font: Self elaborated. SAEB's DATA.

On every analyzed year, the average scores reached by Pernambuco's students are well below the national average and the minimum baseline of knowledge required by the Ministry of Education (200 points for Portuguese and Mathematics). From another point of view, the Figure 7 shows that only 15% of Pernambuco's 4th grade pupils attained the grade's appropriate Mathematics performance.



Font: Self elaborated. SAEB/INEP's data.

Figure 8- A spatial illustration of Student´s Mathematics' score in Pernambuco

[3] The Ministry of Education established a minimum satisfactory baseline of 200 points on SAEB's exams for 4th grade students.

Metodology:Quantil		
Class	**Color**	**Interval**
Classe 0		Less than 166.8756 points
Classe1		166.8756 --| 171.2073 points
Classe2		171.2073 --| 175.6628 points
Classe3		175.6628 --| 180.7161 points
Classe4		180.7161 --| 224.9653 points

Viewing the figure 8, it is possible to realize that the municipalities in Pernambuco are very heterogeneous regarding Mathematics student's performance in SAEB (2007). There is a concentration of worse scores at the hinterlands and at the coast and few were the municipalities which have achieved a 180 to 224 points performance (appropriate for the grade). Not even the capital has reached this point zone, reaching only an average performance of 175 points.

2.2 DATA DECRIPTION

The following article is using the most recent National Basic Education Evaluation System, SAEB, corresponding to the year of 2007, and its respective database in reference to the 4^{th} grade of the Elementary Education. More specifically, it will use the SAEB's Mathematics grades from all Pernambuco state's public and private schools as a variable response or dependent that is wanted to model.

The other variables chosen to be used on the hierarchical regression model in this paper were obtained through questionnaires answered by students/schools and teachers in parallel to the exams, and their inclusion on the estimates is supported by the extensive national and international literatures about the factors that have shown a significant correlation for the determination of students' learning process.

The income and the students' family background characteristics, such as their mother's educational level, and the possession of material goods, are responsible for the socioeconomic control of the model.

The following Table 1 shows the descriptive data of all the variables used on the estimates. It was divided between capital and other counties statistics in order to make explicit the differences and similarities between both regions and for a better visualization and comparison of final results.

Table 1 – Descriptive Analysis

Sample	Schools from capital				Schools out of the capital			
	40.04%				59.96%			
Students Variables	CAPITAL				NON - CAPITAL			
	Mean	std. dev.	Min	Max	Mean	std. dev.	Min	Max
Gender (Male)	0.521	0.500	0	1	0.512	0.500	0	1
Student race	2.120	1.092	1	5	2.035	1.025	1	5
Student has a computer at home	0.426	0.495	0	1	0.259	0.438	0	1
Student has television at home	0.978	0.146	0	1	0.964	0.187	0	1
Student has a car at home	0.432	0.495	0	1	0.341	0.474	0	1
Number of bathrooms at home	1.478	0.753	0	3	1.247	0.589	0	3
Mother schooling	2.410	1.972	0	5	2.114	1.842	0	5
Father schooling	2.141	2.080	0	5	1.729	1.864	0	5
Father read	0.921	0.270	0	1	0.851	0.356	0	1
Flunk	0.252	0.435	0	1	0.349	0.477	0	1
Kindergarten	0.655	0.476	0	1	0.601	0.490	0	1
Student´s age	4.465	1.115	3	6	4.154	1.176	1	6
The student has not made math homework	0.025	0.156	0	1	0.024	0.152	0	1
Class time	436	56	310	730	424	32	330	530
School and Principal Variables	CAPITAL				NON - CAPITAL			
	Mean	std. dev.	Min	Max	Mean	std. dev.	Min	Max
There is no library in the school	0.016	0.124	0	1	0.035	0.184	0	1
has building programs	0.846	0.362	0	1	0.627	0.484	0	1
Principal schooling	0.964	0.186	0	1	0.959	0.197	0	1
Principal salary (gains above R$ 1.700,00)	0.545	0.498	0	1	0.204	0.403	0	1
Public school	0.545	0.498	0	1	0.602	0.490	0	1
Principal selected	0.210	0.143	0	1	0.055	0.229	0	1
Turnover of teachers	0.298	0.457	0	1	0.181	0.385	0	1

Source: SAEB 2007

Notes: mother and father schooling: 0-the pupil doesn´t know about mother´s schooling; 1-none years of school; 2-doesn´s finished elementary school; 3- finished elementary school;4- high school; 5-university

Student race: 1-white; 2-brown; 3-black; 4-yellow; 5-indian

Student´s age: 1-8 years old; 2-9 years old; 3-10 years old; 4-11 years old; 5-12 years old; 6-13 years old. Principal schooling: 0-without university graduation; 1-with university graduation

Number of bathrooms at home: 0-There's no bathroom at home;1-there´s one;2-there´s thwo;3-there´s three bathrooms at student´s home.

The data reveal, through the variables which represent the ownership of material goods, that students who study outside of the capital are poorer than those who live in the capital, as already expected. By the statistics of mother´s and father´s schooling, it is also possible to conclude that the capital children´s parents, on average, have a higher level of education. Some characteristics of the pupil´s academic history, such as kindergarten and flunking, show again the disadvantage of the children from outside the capital. Finally, some schools' characteristics indicate reasonable differences between the two regions. An example is the principal´s salary, which 54% of them reach more than R$ 1.700,00 per month in the capital, meanwhile, in the other cities of the estate only 20% of the principals gain above this sum.

RESULTS ANALYZES

The estimation of model (2) is described in table 2. The standard deviation of schools' random intercept ($\sqrt{\varphi}$) is 26,6, and students' random intercept standard deviation ($\sqrt{\theta}$) is 36. Comparing the variance estimation at tree model (table 2) both variances are reducing as soon as we are including the level-2 and level-1 variables at null model. For intraclass correlation (ρ), we see that the intraclass correlation for the null model without covariates is estimated as 0.34 for the data set. This was reduced to a conditional intraclass correlation of 0.17 when level-2 covariates are added and to 0.12 when all remaining covariates are added. These results suggested that in Pernambuco Schools are responsible for around 11% of differences in between student`s Math tests.

Table 2 – Null Model

	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
Constant	191.55	0.000	202.9	0.00	157.2	0.000
Kindergarten					8.24	0.000
the student work					-9.64	0.000
Gender (man=1)					6.92	0.000
Race*					-0.29	0.866
Student age					-4.87	0.000
# of car					4.07	0.046
# of Computer					1.10	0.383
# of Bathroom					1.76	0.242
Mother´s schooling					0.36	0.503
Father´s schooling					-0.64	0.226
Father no illiterate					4.21	0.109
Mather no illiterate					0.83	0.798
Math homework					-5.50	0.001
Class-time					0.10	0.006
Public school			-36.29	0.00	-22.15	0.000
Capital			9.17	0.01	8.82	0.004

Principal schooling		-1.98	0.49	0.11	0.964
Principal Salary		3.22	0.00	3.34	0.000
Principal selection		4.85	0.25	-0.89	0.816
Professor rotativit		-6.13	0.16	-10.59	0.007
Library		-2.70	0.08	-2.64	0.048
Randon Effect					
$\sqrt{\varphi}$	26.695	16.742		13.132	
$\sqrt{\theta}$	36.806	36.323		35.335	
Derived Estimation					
ρ	**0.344**	**0.175**		**0.121**	

Source: Self elaborated. SAEB 2007. Note: *race: 1-white;0-others

This estimate represents either a comparison between children of different schools (between effect - $\hat{\beta}_B$), or a comparison between children of the same schools (within effect- $\hat{\beta}_F$). To get the purely between- school comparison and within-school comparison it is necessary estimating both.

The random intercept model implicitly assumes that the between and within effects of the set of covariates that vary both between and within schools are identical since the both have the same regression coefficient. However the estimated between effect may differ from the estimated within effect due to omitted school specific explanatory variables that affect both $\bar{x}_{2.j}$ and school specific residual ζ_j. Probably, schools who children don't have done kindergarten or have a parents with low years of education are using to studying in public or chipper schools. These variables adversely affect student score and could have not been adequately controlled for, so that the between effect is likely to be overestimate of the true effect (in absolute value). We thus have cluster-level confounding omitted variable bias and endogeneity problem.

To deal with this problem we now relax the assumption that between and within effect are the same, by estimating the model (8) and (9) at previous section and test if $\hat{\beta}_B = \hat{\beta}_F$ (table 3).

Table 3 – Testing the null hypothesis

	Variables	Coef.	P>|z|		Variables	Coef.	P>|z|
β0	Constant	154.4	0.00	β15	m_Kindergarten	14.10	0.06
β1	Kindergarten	7.51	0.00	β16	m_Gender	-3.35	0.75
β2	the student work	-8.21	0.00	β17	m_Student age	-1.68	0.58
β3	Gender (men=1)	6.51	0.00	β18	m_# of car	3.04	0.79
β4	Race*	-0.61	0.72	β19	m_Class-time	0.14	0.33
β5	Student age	-4.58	0.00	β20	m_# of Computer	0.58	0.93
β6	# of car	3.58	0.08	β21	m_Race	1.37	0.88
β7	Class-time	-0.08	0.58	β22	m_Mother´s schooling	-2.78	0.36
β8	# of Computer	-0.04	0.97	β23	m_Father´s schooling	2.61	0.37
β9	# of Bathroom	-0.78	0.61	β24	m_Father no illiterate	-15.18	0.40
β10	Mother´s schooling	0.44	0.42	β25	m_Mather no illiterate	0.11	1.00
β11	Father´s schooling	-0.87	0.11	β26	m_Math homework	-7.04	0.44
β12	Father no illiterate	5.04	0.06	β27	m_work	-20.89	0.09
β13	Mather no illiterate	1.04	0.75	β28	m_# of Bathroom	29.35	0.00
β14	Math homework	-4.72	0.01	β29	Public school	-3.33	0.42
				β30	Capital	6.48	0.01
				β31	Principal schooling	-0.74	0.72
				β32	Principal selection	0.47	0.88
				β33	Principal Salary	2.01	0.01
				β34	Library	-1.95	0.07
				β35	Professor rotativit	-6.77	0.03
		Chi2	P>|Chi2|				
β15 = β18 = … = β27 = 0		94.39	0				
β15 = β18 = … = β26 = 0		12.47	0.4893				
$\sqrt{\varphi}$		8.248	1.166				
$\sqrt{\theta}$		35.3	0.556				
ρ		0 .051	0.014				

Source: Self elaborated. SAEB 2007

We can formally test the null hypothesis that corresponding coefficients are the same, H0: $\hat{\beta}_F = \hat{\beta}_B$. The results suggested there isn't evidence that both effect are different. We know that $x_{2ij} - \bar{x}_{2ij}$ is an instrumental variable for x_{2ij} because it is correlated with x_{2ij} but not with ζ_j. However, ζ_j could be correlated with another within school covariate. To address this problem we can follow MUNDLAK (1978) and include the cluster means of all with in school covariates. The Wald statistics 94.39 (df = 13) and 12.47 (df=12), suggested that null hypothesis the coefficients of cluster mean are all zero is rejected at the 5% level, but the null hypothesis the coefficients of cluster mean are all less one zero is not rejected at the 5% level. This null hypothesis is equivalent to the hypothesis that between and within effect are the same. We cannot reject this hypothesis for all covariates except number of bathroom.

Instead of including cluster means of all level-1 covariates, we may thus for instance proceed by including cluster means for only the particular covariates where the within and between effects are significantly different at the 5% level and to test the null hypothesis that $\hat{\beta}_R$ is the most efficient model using the Hausman endogeneity statistic test.

There is evidence for the random model is a correct specification since Hausman test statistic is 5.90 with df = 16 we couldn't reject the null hypotheses at the 5% level. Finally, we will use the random model including bath mean as covariates and analyze the impact of fixed effect and the school effect in the math student score in the capital and outside it.

Table 4 - Random Coefficient Model to compare capital and non capital Math Students Score.

	State		Capital		Non Capital	
Fixed Effect	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
Constant	134.83	0.00	119.16	0.00	151.52	0.00
Kindergarten	8.25	0.00	10.23	0.00	6.92	0.00
student works	-9.14	0.00	-9.49	0.04	-8.88	0.00
Gender (1=man)	6.60	0.00	5.37	0.05	7.60	0.00
Race*	-0.47	0.78	1.14	0.70	-1.78	0.38
Student age	-4.56	0.00	-4.03	0.01	-4.91	0.00
# of car	3.53	0.08	-0.07	0.98	6.46	0.01
# of Computer	-0.05	0.97	2.09	0.30	-1.82	0.25
# of Bathroom	-0.71	0.64	0.48	0.85	-2.01	0.28
m_# of Bathroom	35.00	0.00	18.79	0.01	43.12	0.00
Mother´s schooling	0.38	0.49	0.75	0.42	0.08	0.91
Father´s schooling	-0.78	0.14	-1.36	0.12	-0.33	0.62
Father no illiterate	4.70	0.07	14.40	0.01	1.00	0.73
Mather no illiterate	1.05	0.75	0.66	0.92	0.86	0.81
No Math homework	-5.11	0.00	-6.92	0.01	-3.98	0.05
capital	6.37	0.01				
Public school	-5.36	0.14	-17.11	0.02	-2.84	0.47
Class-time	0.05	0.08	0.07	0.14	0.02	0.67
	-1.17	0.56	3.57	0.29	-2.23	0.36
Principal selection	0.25	0.93	7.61	0.16	-3.30	0.34
Principal Salary	1.48	0.04	3.89	0.01	0.65	0.40
No Library	-1.87	0.08	-2.89	0.13	-1.40	0.24
Professor's turnover	-7.43	0.02	-7.93	0.14	-6.76	0.07
Randon Effect						
$\sqrt{\varphi}$	0.089		0.090		0.068	
$\sqrt{\theta}$	0.353		0.369		0.341	
Derived Estimation						
	0.050		5,6%		3,8%	

Source: Self elaborated. SAEB 2007. Notes: *race:1-white ;0-others; p-values in parenthesis

The literature shows evidence that kindergarten exercises long-term influence on individual's schooling. (BERLINSK et al., 2006; FEINSTEIN et al., 1998). Indeed, the result above follows this evidence, and shows a higher score for those children who studied in kindergarten (see CURI and MENEZES, 2006; CAFIERO et al, 2007).

Reflection of dropout, failed and late entry into school, the age-grade discrepancy, obtained by the variable "age", is highly correlated with worse scores. As brought about by SOARES and MENDONÇA (2003), BARBOSA and FERNANDES (2001), the student's performance is heavily penalized for each school year delay.

In relation to the gender variable, which always presents positive significance when it is the male gender, does not deviate from previous evidences, but the results found here suggest that the boys and girls are more homogenized in de big cities (in terms of score) than those from small and medium schools.

Most of the model's, covariables which represent the living standards and the student's social situation, say, the race, the presence of a computer and the level of education of the student's mother and father, did not turn themselves to be significant, possibly because their parcels of influence are already being diminished by two other socioeconomic estimated variables: "father iliterate" and average number of bathrooms at the house of each child per school (mn_bathroom). The first one is particularly significant in the case of capital´s schools and the second one seems, in fact, to represent the influence of family income on the student's score, which reaches a level of significance of 1% in schools outside the capital.

Another fairly relevant information, and often cited by the literature [1], is doing homework. The estimates of this article confirm the empirical evidence of the positive impact of doing the Mathematics task on the student's score. If the student often does not do his/her Mathematics tasks the estimates show a decrease from up to 5 points on the student's score (at the capital), which immediately suggests the importance of parental encouragement of the children's school responsibilities, and, moreover, the incentive for teachers to give and collect their students' homework.

When it comes only to public schools, what can found in the capital is a significantly weaker schools' performance when compared to private schools. On the other hand, in cities outside the capital, this differential is not statistical significant; in other words, the public and private schools from outside the capital have a more homogenized learning than capital's schools. This is quite intuitive, since the capital is the largest concentration of high level of education private schools.

Regarding the variables belonging to the level 2 of the hierarchical model, say, the school variables, the principal’s salary appears significant only at capital schools. At the same way, the principal having passed through selection process for his/her acceptance at the referred school doesn´t seems to make a considerable difference in both capital and non capital cities. Regarding to school’s infrastructural factor, represented by the variable "No library”, the estimate shows that the absence of this space in schools, impacts negatively on the Mathematics grades at the state as a whole.

Now it is possible to answer the question asked at the beginning of this article, i.e., results indicate heterogeneity between schools’ influence on student´s performance. The children who study at the capital have a significantly better performance in SAEB’s Mathematics exam than children who study at other cites. It is well known that the capital’s (Recife) holds state’s largest GDP, and that this index is generally a good measure of the socioeconomic development rate of the city (MACHADO et al, 2008). It is no coincidence that the most economically favored region, in this case, the capital, has superior performance to other cities, suggesting a Cause X Effect relationship between the region’s richness and a better resourcefulness in schools, converging with other results already present in the literature, such as the paper of Machado et al (2008). The author encountered, through hierarchical models, 4th grade students of the Mining Triangle and Vale do Rio Doce (region with the largest GDP of the state) had a superior performance in comparison to those students of metropolitan region of Belo Horizonte, and concludes that this result is due to the favorable economic conditions (logarithm of GDP per capita) of municipalities belonging to both regions with better grades.

Lastly, the school's contribution on student learning evidence is analyzed, represented in this article by the coefficient of correlation between schools. For the state of Pernambuco, it is verified that the share of school’s responsibility on the students’ Mathematics performance revolves around 5%. This value shows that even after socioeconomic differences control between the students from various schools, most of the proficiency variation must still be assigned to students’ intrinsic variations. However, the remaining value is large enough to identify that there is variation between schools, so that the attended school makes difference in student’s life (SOARES 2004).

It is worth of notice that this result is consonant not only with national literature, but also with the international. For the Brazilian case, authors such as FERRAO et al. (2003), BARBOSA and FERNANDES (2001) and FELÍCIO and FERNANDES (2005) found school's effect that ranged from 8% to 19% in Mathematics. Without major differences, the highlights international bibliographic review of Tedlie and Reynolds, 2000 (apud Andrade and Soares 2008) shows an effect school ranging from 11% and 12% in the Netherlands, 11% in Germany and 10% in the USA.

Finally, when compared the heterogeneity of the school´s quality between capital and the other cities of Pernambuco. We realize that the schools from the capital are more heterogeneous, presenting a ρ (beween effect) of 5,6% against only 3,8% from the other municipalities of the state. Although small, this difference shows that there are higher quality schools at the capital, in other words, schools which enable students to achieve a higher academic performance than expected, given their social background. This may be a result, among other factors, of the regional disparity that exists between the capital, Recife, and other cities in the state (which are less economically developed), thus being considered a measurement of favorable socioeconomic conditions of the city acting on the quality of education.

5. CONCLUSION

One student's cognitive performance, measured by his/her standardized tests proficiency, is a result of personal choices, of socioeconomic antecedents, of family structure and values, of the society in which he/she lives in, and, finally, the school in which he/she studies (ANDRADE E SOARES, 2008). This last factor, for being more willing to public policies interventions, has become target of investigations and studies that aim to explain the school`s parcel of influence on the student`s performance after socioeconomic control. Building through this line of research, this present article intended to estimate the Pernambuco`s 4^{th} grade school effect and compare the results attained between capital and non-capital schools.

Is well known the complex interaction between the innumerable factors that act simultaneously in many student's social insertion levels. For this, it is convenient the effect that the school has should be isolated and measured with the usage of statistic models which control the influence of other determinants. With effect, this paper uses the hierarchical regression models to estimate the school's positive contribution, considering the 2007 SAEB`s Mathematics score as variable response.

The results reveal that capital students attained a better performance in mathematics tests in comparison to non-capital students. Such evidence is treated, in a roughly way, as measurement of favorable socioeconomic conditions acting positively on teaching quality, since the capital (Recife) is richer and more productive than other cities of the state.

In relation to the variables that were inserted in the regression model; pre-school, homework, library in school, and selected principal must be highlighted as factors that carry a positive and statistic relevant impact in reference of Mathematics grades. On the opposite way, student`s academic history characteristics, such as age grade distortion and flunking, were revealed, as expected, as being quite harmful, specially to those non-capital students.

Finally, the obtained school effect, from 5,6% at the capital to 3,8% outside it, is in harmony with the literature, being very intuitive to discern that this effect is a little bit larger in the state`s richest region.

Although being a small number, it is expressive enough to conclude that the school is a differential and that such values may serve as reference to government actions for the educational policies to work the potential and the social function of school in promoting the social leverage, dreduce regional disparities and revert the poverty vicious cycle.

6. BIBLIOGRAPHY

BARBOSA, M. E. F.; FERNANDES, C. A Escola brasileira faz diferença? uma investigação dos efeitos da escola na proficiência em matemática dos alunos da 4ª série.

BARROS, R.P. e REIS, J.G.A. Desigualdade Salarial e Distribuição de Educação: A evolução das Diferenças Regionais no Brasil, Pesquisa e Planejamento Econômico, v. 20, n.3,1990.

CURI, A., MENEZES-FILHO, N., Os efeitos da pré-escola sobre os salários, a escolaridade e a proficiência escolar. Anais da ANPEC 2006.

FELÍCIO, F. e FERNANDES, R.. O Efeito da Qualidade da Escola sobre o Desempenho Escolar: uma Avaliação do Ensino Fundamental no Estado de São Paulo, Anais do XXXIII Encontro Nacional de Economia, 2005

FERRÃO, M. E. et al. O SAEB – Sistema Nacional de Avaliação da Educação Básica: objetivos, características e contribuições na investigação da escola eficaz. Revista Brasileira de Estudos de População, v. 18, n. 1/2, 2001.

HANUSHEK, E. A. School resources and student performance. Does money matter? The effect of school resources on student achievement and adult success. Washington: The Brooking Institution, 1996.

HANUSHEK, E. A. e LUQUE, J. A. Efficiency and Equity in Schools Around the World. Economics of Education Review, v. 5, n. 22, 2002

HANUSHEK, E. A. e RAYMOND, M. E. School Accountability and Student Performance. Federal Reserve Bank of St. Louis, Regional Economic Development, v.2, no 1, 2006.

MAMBRINI, J., SOARES, J.F., CÉSAR, C.C. Determinantes de desempenho dos alunos do ensino básico brasileiro: evidências do SAEB 1997. Relatório Interno. Departamento de Estatística da Universidade Federal de Minas Gerais, 1999.

MENEZES FILHO, N. Os Determinantes do Desempenho Escolar no Brasil. Instituto Futuro Brasil, São Paulo, Ibmec – SP, 2007

MORTIMORE, P. The nature and findings of school effectiveness research in primary sector. In: RIDDELL, S.; PECK, E. (Org.). School effectiveness research: its message for school improvement. Londres: HMSO, 1991.

PERNAMBUCO, SECRETARIA ESTADUAL DE EDUCAÇÃO. Educação em Direitos Humanos como política de Estado: educando na diversidade e na diferença, 2007.

RAUDENBUSH, S. W.; BRYK, A. S. Hierarchical linear models. Thousand Oaks: Sage, 2002.

RIVKIN, G.; HANUSHEK, E . e KAIN, F. Teachers, Schools, and Academic Achievement, Econometrica. vol.73, n..2, 2005.

SAMMONS, P.; HILLMAN, J.; MORTIMORE, P. Key characteristics of effective schools: a review of school effectiveness research. London: Office for Standards in Education [OFSTED], 1995.

SOARES, J. e ANDRADE, R., O efeito da escola básica brasileira. Estudos em Avaliação Educacional v.19 n. 41, 2008

SOARES, J. O efeito da escola no desempenho cognitivo de seus alunos, Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación,Vol. 2, N. 2, 2004.

WILLMS, J. D.; SOMERS, M-A. *Schooling outcomes in Latin America*. A report for UNESCO, 1999