Escola de ECONOMIA de São Paulo

Textos para Discussão

282

Maio de 2011

C-Micro Working Paper Series

02

Maio de 2011

Changes in Test Scores Distribution for Students of the Fourth Grade in Brazil: A Relative Distribution Analysis for the Years 1997 to 2005

Clarissa G. Rodrigues
Eduardo L. G. Rios-Neto
Cristine C. X. Pinto

FGV EESP

Changes in Test Scores Distribution for Students of the Fourth Grade in Brazil: A Relative Distribution Analysis for the Years 1997 to 2005

Clarissa G. Rodrigues (CEDEPLAR/UFMG); Eduardo L. G. Rios-Neto (CEDEPLAR/UFMG); Cristine C. X. Pinto (EESP/FGV)

Abstract

To assess the quality of school education, much of educational research is concerned with comparisons of test scores means or medians. In this paper, we shift this focus and explore test scores data by addressing some often neglected questions. In the case of Brazil, the mean of test scores in Math for students of the fourth grade has declined approximately 0,2 standard deviation in the late 1990s. But what about changes in the distribution of scores? It is unclear whether the decline was caused by deterioration in student performance in upper and/or lower tails of the distribution. To answer this question, we propose the use of the relative distribution method developed by Handcock and Morris (1999). The advantage of this methodology is that it compares two distributions of test scores data through a single distribution and synthesizes all the differences between them. Moreover, it is possible to decompose the total difference between two distributions in a level effect (changes in median) and shape effect (changes in shape of the distribution). We find that the decline of average-test scores is mainly caused by a worsening in the position of all students throughout the distribution of scores and is not only specific to any quantile of distribution.

Keywords: test scores; relative distribution; Brazilian education

1. Introduction

The evaluation of the quality of elementary school education, measured by students' performance on standardized tests, has been the focus of many studies in educational and economic literature in recent years due to the existence of many standardized tests that evaluate students in all levels of education. For example, the PISA (Program for International Student Assessment) and the TIMSS (Trends in International Mathematics and Science Study) can be considered as the two major tests that attempt to measure students' math and science performance across countries. This marked a major advancement in measuring the quality of education, because the measurements available thus far, such as the number of years in school, are limited to soley capturing the quantitative dimension of education (Hanushek, 2005).

In general, data on students' test scores is explored in comparative analyses that seek to understand why the national averages of certain countries are higher than in others (Ammermüler, 2004; Wang, 2004; Woessmann, 2004; Carnoy, 2007) or in case studies what the goal is in finding out how the individual and family background variables and class and school characteristics are capable of increasing the average cognitive achievement of students (Behrman *et al.*, 1997; Fertig, 2003). The inherent objective in all of these studies is to know the characteristics that positively relate to students' school performances with the intention of supplying input for the formulation of public policies imbued with the task of improving the quality of education.

A significant part of the educational research is concerned with comparisons of test scores means or medians. The research uses traditional statistical methods based on linear regression and its extensions, like hierarchical linear models, which are not designed to represent the rich detail of distributional patterns of the data. They focus on the modeling of conditional means and, as a result, the research leaves most of the distributional information in the data untapped. The analysis of full distribution is important because students at different parts of the distribution couldn't have the same return in education as students in the mean/median. At the same time, the analysis of full distribution allows us to explore some aspects of educational quality and inequality.

In this paper, we shift the focus on means/medians and explore test scores data by addressing some often neglected questions. For example, in the case of Brazil, the test scores average declined at the end of the 1990s. But what about changes in the distribution of scores? It is unclear whether the average decline was caused by deterioration in student performance in upper and/or lower tails of the distribution. If the decline of the average is the result of a worsening in performances of the students of a lesser capacity, then we are facing a scenario of a fall in average quality and a rise in educational inequality. However, if the decline of the average is a consequence of a deterioration in the performance of the students of a greater capacity, then the scenario is that of a reduction in inequality, but in the lower levels of elementary school performance.

The objective of this study is to explore the decline in average elementary school performance of Brazilian students with a distributional perspective. To deal with this little explored aspect in the literature, we use the relative distribution method developed by Handcock and Morris (1999). This method is a statistical tool for fully representing differences between distributions. It was developed to analyze income inequalities in

the labor market and it will be applied to educational data in this paper. The advantage of this methodology is that it lets us compare two distributions of test scores through a single distribution and synthesizes all the differences between them. Moreover, it is possible to decompose the total difference between two distributions in a location effect (changes in median) and shape effect (changes in the shape of the distribution).

We use data from the Brazilian National System for Evaluation of Basic Education (SAEB) for the period between 1997 and 2005. SAEB collects data about the cognitive achievement of student enrollment in the fourth and eighth grades of elementary school and third year of secondary school for Math and Portuguese language. Moreover, SAEB collects information about the home and school environment, such as parents' education, school infrastructure, the quality of teachers and principals, etc. In this paper we use data for a cohort of fourth graders evaluated in math test scores.

Following this introduction, this paper has six sections. Section 2 presents a descriptive analysis of the mean and distribution of the test scores for the students of the fourth grade in Brazil. Section 3 describes the Brazilian data used in this paper and the descriptive analysis. Section 4 describes the empirical strategy based on the relative distribution method. Section 5 presents and analyses the results. Finally, section 6 presents the final remarks.

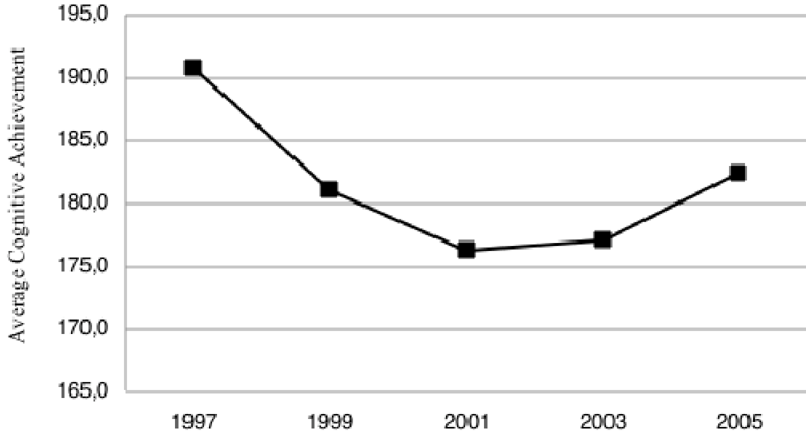
2. Why study the entire distribution of test scores?

The Brazilian system of basic education produced poor results with respect to quality of education between 1997 and 2005. An analysis of the evolution of the average elementary school performance of 4th grade students in Brazil[1] can be seen in Graph 1. It can be observed that over the years average elementary school performance fell, exacerbating the problem of low quality of education in Brazil. Even in 1997, where the largest average was within this historic series was obtained, it is noticeable that students' learning was deficient, since it was below the level considered adequate for the grade in question[2]. If the students do not have command of the minimum prerequisites for the fourth grade, they will continue on this path with deficiencies that further compromise their learning.

Although analyses based on simple summary measures are often very informative, they do not always tell the whole story. Hidden behind these summary statistics are a range of important questions. Given that the average is a summary measurement of the distribution of the scores that measure the students' learning, its alteration is associated with changes of the scores in specific segments of the distribution. It is important to know these aspects since a decline in the averages can have different implications on education policy depending on the segment of students whose elementary school performance was affected.

[1] Brazil organizes its education system into five levels: infant, elementary, secondary, higher and postgraduate education. The first three are referred to as basic education.
[2] According to INEP, students need a performance above 250 points in math to be compatible with 4th grade expectations. Students above this level are capable of competently doing the four arithmetic operations (addition, subtraction, division and multiplication), besides recognizing elements and characteristics peculiar to geometric planes.

Graph 1
Average-test scores in Math, 4th grade of elementary school, Brazil, 1997 to 2005



Source: INEP, National System for Evaluation of Basic Education (SAEB), 1997, 1999, 2001, 2003 and 2005.

In Brazil, the negative evolution of elementary school performance averages was attributed, by researchers and government officials, to the process of expanding basic education and the non-repetition policy (Souza, 2006; Neri and Carvalho, 2002; Fernandes and Natenzon, 2003; Alves, 2007). The rise in elementary school enrollment occurred between children and youth from a low social background and, consequently, with greater learning difficulties. The non-repetition policy[3], for its part, incentivized the permanence of students in schools, reducing the dropout rates of those students with low levels of learning. As a result, these policies allowed for an increase in the proportion of less qualified students with regard to cognitive abilities, which would explain the reduction in the average-test score over the years.

Comparisons of average test scores without accounting for the changes in the entire distribution of achievement can generate a misleading inference about the relationship between the educational expansion and variation of school outcome. If the reduction of the averages is due to changes in the performance of students situated at the base of the distribution who, on average, are from low income backgrounds, it is necessary to think of reinforcement/tutoring policies that could minimize the deficiencies in learning. This reduction in performance generally stems from the lack of support or preparedness of the parents regarding help with homework, the lack of a healthy environment propitious to learning in students' homes, the absence of a dialogue between parents and children regarding the importance of education on improving future life conditions and other various deficiencies associated with the low income and schooling of parents.

On the other hand, if the reduction of the averages is also due to changes in the performance of students situated at the top of the distribution who, on average, have a better family background, it is necessary to know the origin of this problem. These students have pecuniary and non-pecuniary incentives that allow for an improved level of education, which, initially, would lead us to believe that their poor performance could be related to a problem outside the family context and, therefore, connected with the effectiveness of the education system.

[3] The principal objective of this policy is to guarantee that students remain in school by eliminating repeat years. The debate about this type of policy can be seen in N'tchougan-Sonou (2001).

In recent years, international literature has presented studies not only focused on the "average" student, but also on the entire distribution of achievement. The subjacent idea in these studies is that the effects of school and the family context on school results can vary depending on the segment of the population of the students that are being analyzed, i.e., depending on the level of qualification of the student. Levin (2001) evaluates the impact of class size and peer effects on the distribution of math and language achievement for Dutch primary school students; Corak and Lazon (2002) apply semi-parametric decomposition based on the changes in full distribution to assess the contribution of differences in family background and school characteristics to differences in achievement distributions between Canadian provinces; Fertig (2003) looks for determinants of German student's achievement, like family background information and the characteristics of school and class, based on quantile regressions; Ding and Lehrer (2008) assess the benefits of the reduction of class size for the lowest and highest performing students based on the results of Project STAR; among others.

In the national literature, Soares (2006) calls attention to the importance of analyzing the distribution of cognitive achievement as a whole in order to evaluate, in tandem, aspects relating to educational quality and equity. In the research, Soares (2006) develops a methodology to measure education inequality based on data related to students' cognitive achievement. As a starting point, the author uses the idea subjacent to the Gini index, an instrument used to measure income inequality, and makes some methodological modifications to show that the measurement of education inequality differs from the measurement of income inequality. Education inequality is measured by the distance between an ideal distribution, given by the distribution of the students assessed by PISA, and the empirical distribution of school performance of students assessed by SAEB in Brazil. In Brazilian literature, we can also cite the study by Pinto (2010) which evaluates the existence of peer effect on the entire distribution of math and language achievement for the students in the fourth grade of elementary school in Brazil.

Considering the relevance of the analysis of school performance as a distributional perspective and the lack of studies along this line where Brazil is concerned, we seek, in this paper, to explore this point further. In the following section, we present the data base, some descriptive statistics and, subsequently, the methodology used to discover some neglected aspects in the literature that use students' test scores to evaluate the quality of the educational system.

3. Data and descriptive statistics

3.1 Data

In this section we describe the Brazilian data set used in our analysis. We use the National System for Evaluation of Basic Education (SAEB), which is a biennial survey conducted by the National Institute of Educational Studies and Research (INEP), an agency of the Ministry of Education (MEC). SAEB evaluates the cognitive process of students currently enrolled in the fourth and eighth grades of elementary school and in the third year of secondary school in two subjects: Math and Portuguese language.

The level of cognitive achievement of each student is calculated through a test based on the National Curricular Parameters (PCN) and the National Educational Guidelines and Framework Law (LDB), and the results of a thorough national consultation with teachers, researchers and specialists. SAEB began in 1990, but only since 1995 can the results be compared between years and grades because of the introduction of a new methodology, the Item Response Theory, to compute students' proficiencies. The scale of SAEB ranges between 0 and 500. The higher value in scale means better student performance.

In our research, we are especially concerned with the period between 1997 and 2005 which include a higher decline in average test scores. The sample used in this paper and the descriptive statistics are presented in Table 1. The sample includes students enrolled in public (except federal schools) and private schools, all of them located in urban areas.

Table 1
Sample and descriptive statistics, fourth grade of elementary school, Math, Brazil, 1997 to 2005

Year of SAEB	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
1997	18.588	192,29	43,86	80,37	377,95
1999	16.811	182,17	41,00	82,72	355,93
2001	50.782	179,17	46,01	59,84	367,25
2003	40.596	180,68	45,01	66,42	369,98
2005	37.719	186,09	47,18	65,43	373,44

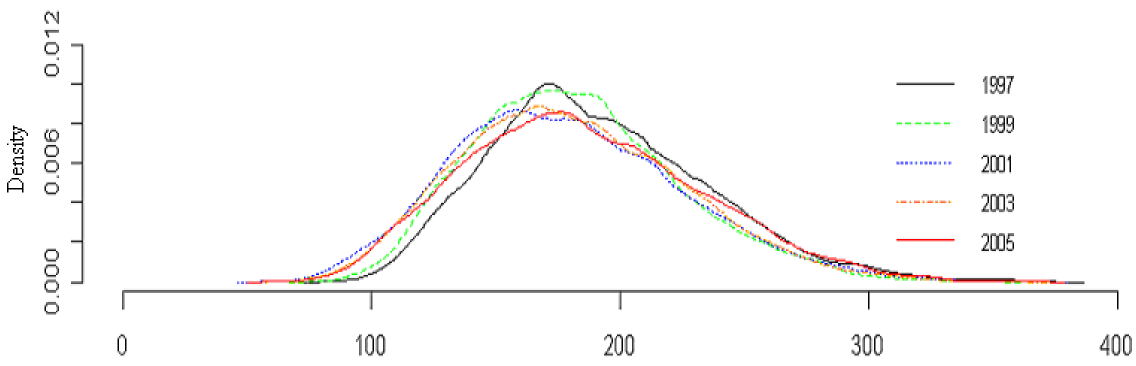
Source: INEP, National System for Evaluation of Basic Education (SAEB), 1997, 1999, 2001, 2003 e 2005.

3.2 Descriptive analysis of test scores distribution

Before embarking on a more complex statistical analysis of changes in the distribution of school performance, a descriptive analysis of the data over the years is needed. To see how the empirical distribution of school performance behaved between 1997 and 2005, we present, in Graph 2, the probability densities for each year. In comparing the curves, the first aspect observed is the leftward shift of the probability density curves of school proficiency of the students evaluated after the year 1997. Besides this shift, these curves are noteworthy for having a larger number of students with lower cognitive abilities, as can be seen by the fat left tail. The distribution in 2001 reaches the apex of this expansion. Also, it is possible to note, a slight straightening of the superior tail, mainly in 1999, indicating a reduction in the density of students with a higher performance. Therefore, apparently a change in the location and dispersion of the distribution exists.

Graph 2

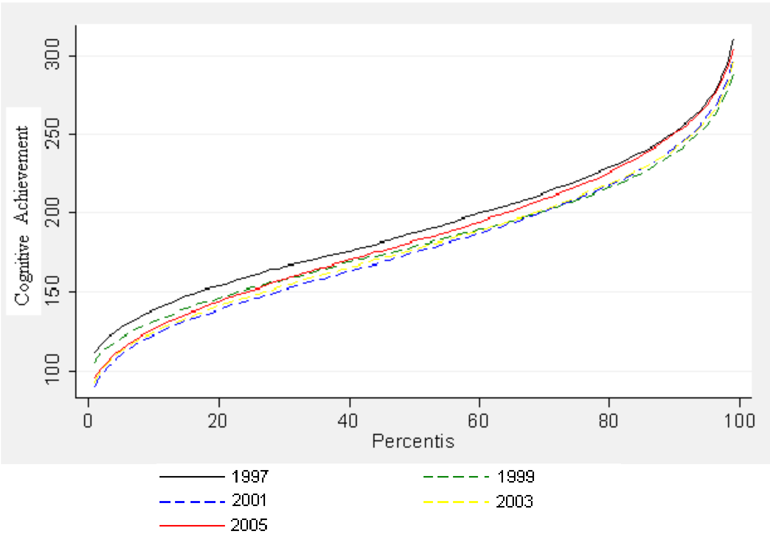
Probability density function of test scores, fourth grade of elementary school, Math, Brazil, 1997 to 2005



Source: INEP, National System for Evaluation of Basic Education (SAEB), 1997, 1999, 2001, 2003 ~~e~~ and 2005.
Note: The densities were estimated by Kernel Epanechnikov and the bandwidth was calculated by the rule of thumb of Silverman (1986).

In the case of a change in location, the set of students as a whole would present a reduction in the scores obtained in the proficiency tests. Graph 3 compares the figures of the percentiles of the distribution of the performance in all the periods and Table 2 supplies the figures for the deciles. In analyzing Graph 3 it is possible to notice a reduction in the levels of performance in each percentile (obviously, as well as in the deciles, whose figures are in Table 2) given by the downward shift of the percentile curve. This effect was more intense in the first centiles of the distribution. In 2005, the distribution is similar to the 1997 curve in the last centiles of the distribution.

Graph 3
Percentiles of test scores, fourth grade of elementary school, Math, Brazil, 1997 to 2005



Source: INEP, National System for Evaluation of Basic Education (SAEB), 1997, 1999, 2001, 2003 and 2005.

Table 2
Deciles of test scores, fourth grade of elementary school, Math, Brazil, 1997 to 2005

Deciles	Cognitive achievement in Math				
	1997	1999	2001	2003	2005
1º	138,99	131,29	123,15	124,87	127,11
2º	153,92	146,10	138,74	141,21	144,38
3º	166,19	157,61	151,64	154,02	158,21
4º	176,12	168,75	163,03	165,71	170,65
5º	187,28	178,68	175,29	177,18	182,39
6º	200,05	189,54	187,18	189,06	195,09
7º	213,03	200,96	201,14	202,52	209,50
8º	228,74	215,86	217,61	218,64	226,56
9º	250,97	237,80	241,31	241,13	250,82

Source: INEP, National System for Evaluation of Basic Education (SAEB), 1997, 1999, 2001, 2003 and 2005.

The statistics presented here appear to show that the reduction in average school performance through the period from 1997 to 2005, principally in the first years of this series, is the result of a poorer performance by both the lowest and highest performing students on the math proficiency tests. To better understand this dynamic, we applied the relative distribution method, described below, which permits the analysis of the differences between two distributions and decomposes them for changes within the average and shape of the distribution.

4. Relative distribution method

The relative distribution is a non-parametric descriptive statistical tool used to compare two distributions of the same attribute between groups or periods. In this paper, the attribute of the study is the elementary school performance of fourth grade students and the comparison is done between two points in time: one of which is used as the reference (reference) year (t_0) and the other as the comparison year (t_1).

To formalize the technique, let us suppose that, in each one of these years, we have the probability density functions, $f_0(y_{t_0})$ and $f_1(y_{t_1})$, and the cumulative distribution functions, $F_0(y_{t_0})$ and $F_1(y_{t_1})$, where Y corresponds to elementary school performance. Based on these functions, the primary school relative distribution between t_0 and t_1 can be produced by the rescheduling of elementary school performance in t_1 using the cumulative distribution of performance function in t_0:

$$R = F_0(Y_{t_1}) \tag{1}$$

With this rescheduling, we produce the relative data r, continuous in the interval [0,1], which measures the relative position of Y_{t_1} in the distribution of Y_{t_0}, or rather, since the students evaluated in t_1 would be allocated to the distribution elementary school performance in t_0. Let us consider, for example, that a student with a performance of 200 points from the SAEB scale in t_1 = 2005, is in the sixth decile of the performance

distribution of the students evaluated in $t_0 = 1997$. The identification of this position, or rather, the sixth decile, corresponds to the value of the relative data r.

Since r is a random variable, it also has a probability density function, *g(r),* and a cumulative distribution function, *G(r),* given as:

$$G(r) = F(F_0^{-1}(r)) = F(Q_0(r)) \qquad 0 \le r \le 1 \tag{2}$$

Where $Q_0(r)$ is the quantile function of $F(y_{t_0})$ and r represents the quantile.

The probability density function of the relative data, *g(r),* called "relative density" by Handcock and Morris (1999), can be obtained by (2):

$$g(r) = \frac{f_1(Q_0(r))}{f_0(Q_0(r))} \qquad 0 \le r \le 1 \tag{3}$$

In terms of the original scale of the measurement of elementary school performance, the relative density can be expressed as:

$$g(r) = \frac{f_1(y_r)}{f_0(y_r)} \qquad y_r = Q_0(r) \ge 0 \tag{4}$$

Based on equation (4) we can see that the relative density is calculated by means of a ratio of densities: the ratio of the density of students tested in the period of comparison, $f_1(y_r)$, to the density of students tested in the period of reference, $f_0(y_r)$, at a given level of elementary school performance, y_r (elementary school performance related to quantile r of the reference year distribution t_0).

Thus, for each quantile of the elementary school performance distribution of the reference year (t_0) there are three ways of interpreting the results: i) when the relative density is greater than 1, ($g(r) > 1$), we say that there is an over-representation of the students from the period of comparison in relation to students tested in the period of reference; ii) when the relative density is less than 1 ($g(r) < 1$), this relationship is inverse, in other words, there is an under-representation of the students from the period of comparison in relation to the period of reference; and iii) when the relative density is equal to 1, ($g(r) = 1$), the density of students in the periods of reference and comparison is the same for the quantile in question and this indicates that there is a distributional equivalence[4].

[4] The relative distribution was implemented in the R *software* (*R Core Development Team, 2007*) using the "*reldist*" suite. The syntax was produced based on the adaptation of the routine supplied by Handcock and Aldrich (2002). For the calculation of the relative densities, we utilized the standard method used in the computational programming prepared by Handcock and Aldrich (2002), which is based on the smoothening of the local likelihood method (*local likelihood method*)[4]. According to Handcock and Janssen (2002, p.416), the use of the estimator of maximum local likelihood presents better results when compared with the Kernel estimator, since the latter entails the difficulty of

The results produced by this technique simplify the comparison between two probability density curves as the differences between them are synthesized through a single curve formed by relative density rates. Besides this advantage, the relative distribution is decomposable. Therefore, the total differences existent between the elementary school performance distributions in the years t_0 and t_1 can be explained either by the changes that took place in the location of the distribution (location effect) or by changes prevalent in the design of the distribution curve (shape effect).

To formalize this decomposition, it is necessary to create a hypothetical variable, Y_h, whose level of distribution is equal to the level of distribution of the comparison period t_1, but the structure remains the same as that of the period of reference, t_0. For a change in the mean, Y_h is defined as a random variable represented by $Y_h = Y_{t_0} + p$, where $p = \overline{Y}_{t_1} - \overline{Y}_{t_0}$. In this case, $\overline{Y}_{t_0}$ is the median performance in the period t_0 and $\overline{Y}_{t_1}$ is the median performance in the period t_1. With the three variables, Y_{t_0}, Y_{t_1} and Y_h, it is possible to produce two relative distributions that isolate the effects of changes within the level and structure of the distribution. To generalize notation (1), we have:

- Total effect = relative distribution of Y_{t_1} and Y_{t_0} (equivalent to equation 1):

$$R \equiv R_0^1 = F_0(Y_{t_1}) \tag{5}$$

- Location effect = relative distribution of Y_h e Y_{t_0}:

$$R_0^h = F_0(Y_h) \tag{6}$$

- Shape effect = relative distribution of Y_{t_1} e Y_h:

$$R_h^1 = F_h(Y_{t_1}) \tag{7}$$

To generalize notation (4), the total effects, location and shape can be represented as a function of the relative density rates as follows:

$$\frac{f_1(y_r)}{f_0(y_r)} = \frac{f_h(y_r)}{f_0(y_r)} \times \frac{f_1(y_r)}{f_h(y_r)} \tag{8}$$

The left side of the equation represents the total relative density (density), $g_0^1(r)$; the first ratio on the right of the equation represents the relative density from the location

underestimation of the relative densities for values of the relative data (r) near 0 or 1 (border effect). For a comparison of the methods, see Hall and Tao (2002). For a detailed description of the method, see Loader (1996).

effect, $g_0^h(r)$; and the second ratio on the right of the equation represents the relative density from the shape effect, $g_h^1(r)$.

All the results produced by this technique are presented in graphs and are quantified by summary measures, such as the entropy index (entropy index) and the polarization index (polarization index), whose metrics are described in the following sections.

4.1 Entropy index

The entropy index is used to measure the difference between two distributions. Handcock and Morris (1999) suggest the use of the formalization of Kullback-Leibler, because apart from giving a simple interpretation in terms of the relative distribution, it can be decomposable on the effects of level and structure. Formally, this measurement can be expressed as follows:

$$D(F_1;F_0) = \int_{-\infty}^{\infty} \log\left(\frac{f_1(y_r)}{f_0(y_r)}\right) dF(y) = \int_0^1 \log\left(g(r)\right) g(r)\,dr \tag{9}$$

Based on this measurement, the three components of the decomposition are thus:

$$D(F_1;F_0) = D(F_h;F_0) + D(F_1;F_h) \tag{10}$$

Where:

$D(F_1;F_0)$ = total difference between the proficiency distributions Y_0 and Y_1;

$D(F_h;F_0)$ = difference between the distributions caused by alterations in level;

$D(F_1;F_h)$ = difference between the distributions caused by alterations in structure.

The relative magnitude of the second and third components signal the relative contribution of changes in location and shape over the total differences observed between the distributions in the periods $\boldsymbol{t_0}$ and $\boldsymbol{t_1}$.

4.2 Polarization index

The polarization index corresponds to the median absolute deviation of the relative distribution. What characterizes a polarized relative distribution is the format in U of its density. When this happens, it can be said that there was an increase in the proportion of students within the lower and upper tails of the distribution and, therefore, there was an increase in inequality. Thus, this index permits the visualization of what happens in the center and the lower and upper tails of the distribution, which is not possible when only the trend of the average is measured.

For the construction of the polarization index, we consider the relative distribution of $\boldsymbol{Y_0}$ in relation to $\boldsymbol{Y_1}$ given by $R_0^h = F_0(Y_1 - p)$, where $p = Q_1(1/2) - Q_0(1/2)$, in other words, p is equal to the difference between the mean of $\boldsymbol{Y_1}$ and $\boldsymbol{Y_0}$; Q is the quantile function. Since the mean of both distributions are equal, the mean of R_0^h will be ½. Thus, the relative polarization index of the median – *median relative polarization index* – can be defined as:

$$MRP(F_1;F_0) = 4\int_0^1 \left|r - \tfrac{1}{2}\right| g_h^1(r)dr - 1 \tag{11}$$

This index measures the absolute deviation around the median of the relative distribution utilizing differences in structure only. The distance between the relative data r and the center of the distribution, $\left|r - \tfrac{1}{2}\right|$, is weighted by the value of the density in r, $g_h^1(r)$. The index assumes values between -1 and 1. The value 0 (zero) indicates that there are no differences between F_0 and F_1 associated with changes in structure; positive values of the index indicate that there are differences in the format of the curve of the distribution that led to the increase in the polarization of the relative distribution (increase in the relative densities in both tails); negative values represent a smaller polarization, characterized by a convergence of the densities toward the center of the relative distribution. If the difference between F_1 and F_0 is caused by only differences in level, then, $g_h^1(r)$ will be equal to 1, signaling a uniform distribution, and $MRP(F_1;F_0)$ will be equal to zero, signaling that there are no differences between F_1 and F_0 caused by changes in structure.

The median polarization index can be decomposed into two parts, making it possible to test the contributions of changes in the distribution below the median (*lower index*, equation 12) and above the median (*upper index*, equation 13):

$$LRP(F_1;F_0) = 8\int_0^{\frac{1}{2}} \left|r - \tfrac{1}{2}\right| g_h^1(r)dr - 1 \tag{12}$$

$$URP(F_1;F_0) = 8\int_{\frac{1}{2}}^1 \left|r - \tfrac{1}{2}\right| g_h^1(r)dr - 1 \tag{13}$$

These two indices have similar interpretations to the total polarization, they are symmetric and do not vary with the monotonic transformations of the original measurement. The positive values represent greater polarization, which means thickening in the tails of the distribution. The negative values represent a reduction in the polarization, indicating a trend towards a convergence at the center of the

distribution. The inexistence of the polarization in the inferior and/or superior tails is observed when the index is equal to 0 (zero) [5].

5. Results

To compare the distribution of elementary school performance at two points in time and analyze the differences between them, we used 1997 as the reference year and 1999, 2001, 2003 and 2005, one by one, as years of comparison. In using this strategy of fixing 1997 as the reference distribution for all the years, we can temporarily follow changes in the distribution and analyze them in conjunction with the decline of the average-test scores[6]. In addition to composing the first year of the historical series used in this study, 1997 showed the higher value of mean and quantiles of cognitive achievement in relation to the other years, as seen in the descriptive statistics presented in Tables 1 and 2.

The results of the relative distribution can be seen in Figure 1. Each panel in this display represents a component of the change. Panel A shows the overall change by period; panel B shows the effect of the location shift; and panel C shows the effect of the shape shift. On the x axis, we have a rising distribution of the students in the reference year in tenths of their elementary school performance and along the y axis we have the relative density of the test scores observed in the years of comparison in relation to what was observed in the reference year. The bars represent the observed density of the distribution of elementary school students of the comparison year if they were in order on a test scores scale of the reference year. The curve on the graph corresponds to the leveling of the bars.

Initially, we focused on analyzing the results of Panel A, where we can observe the behavior of the smoothened curve formed by the relative densities. We notice that there is a monotonic decline in the relative densities along the quantiles of the distribution. This trend signals changes in the distribution of elementary school performance between 1997 (reference year) and the comparison years, since otherwise the relative densities would be constant and equal to 1 (sketched line) indicating a distributional equivalence.

The differences between distributions are clearly marked by a strong downshift in test scores, i.e., there is an increase in the density of the students with lower performances in the proficiency tests concurrent with the reduction in the density of students with a

[5] Handcock and Morris (1999) stress that the synthesis-measurements obtained through relative distribution are robust in relation to the presence of outliers and to the violation of the hypothesis concerning the form of the distribution. This robustness comes from two properties of the relative distribution: 1) The rescheduling of the distribution in the comparison period, t_1, in relation to the distribution in the reference period, t_0 (the transformation of the original measurements, Y_0 and Y_1, in terms of the position/quantiles within the interval [0,1] moderates the influence of the abnormal values); 2) since it is based on a non-parametric approach, it minimizes the chances of a violation of the hypothesis.

[6] Another analysis could be done trough a comparison in each biennium. In this case, the reference year could be the first year of the biennium and the objective becomes to analyze the short-term changes in the distribution of test scores. A different form of comparison can be seen in Soares' (2006) study, where the author uses PISA distribution as a reference distribution to analyze changes in quality and equity of the Brazilian educational system.

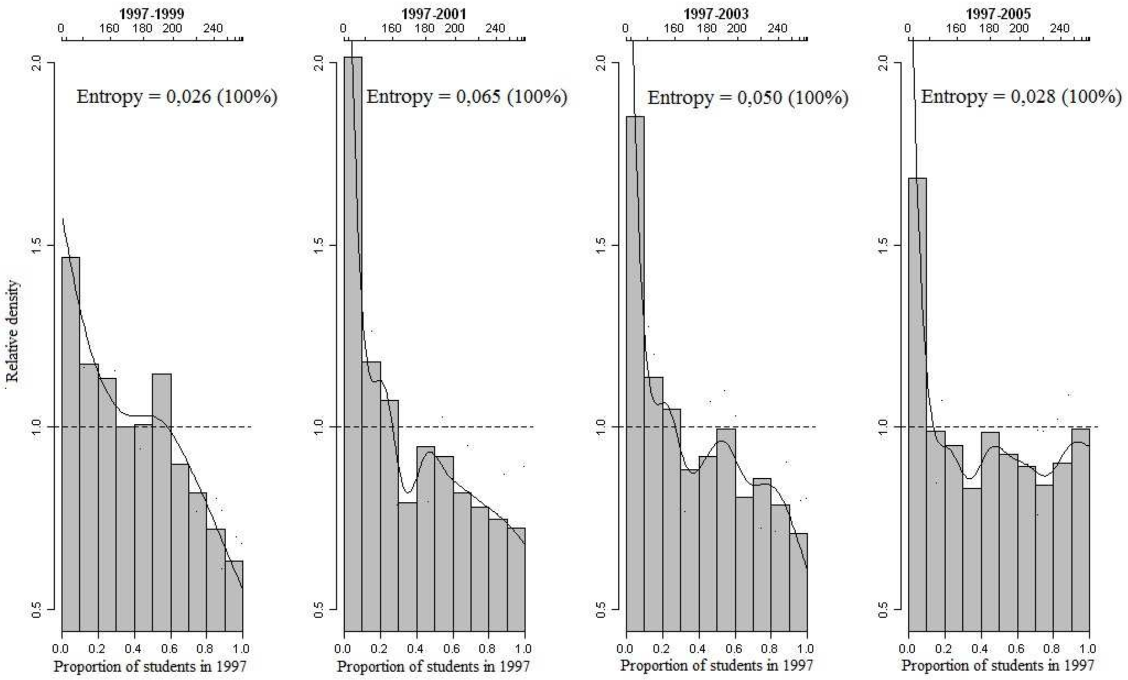
higher performance. For example, between 1997 and 1999, we can observe a relative density of, approximately, g(0.1)=1,5 in the first decile of the distribution. This result indicates a 50 percent increase in the population of students who, in 1999, had a performance compatible with that verified in the first tenth of the distribution in 1997, whose value was 138,85. In the other extreme of the distribution (last decile), we observed a relative density which approximates to g(1.0)=0,6. This result indicates a reduction of about 40 percent in the population of students who, in 1999, had a performance above 250,97 – the level which refers to the last tenth of the 1997 distribution –, when compared to the population of students that in 1997 reached this same level of proficiency.

While the shape of the overall relative distributions clearly points to the location shifts as the dominant trend in each period, it may be masking some of the most subtle changes. To see this, we can focus on panel B (location shift) and C (shape shift) in Figure 1.

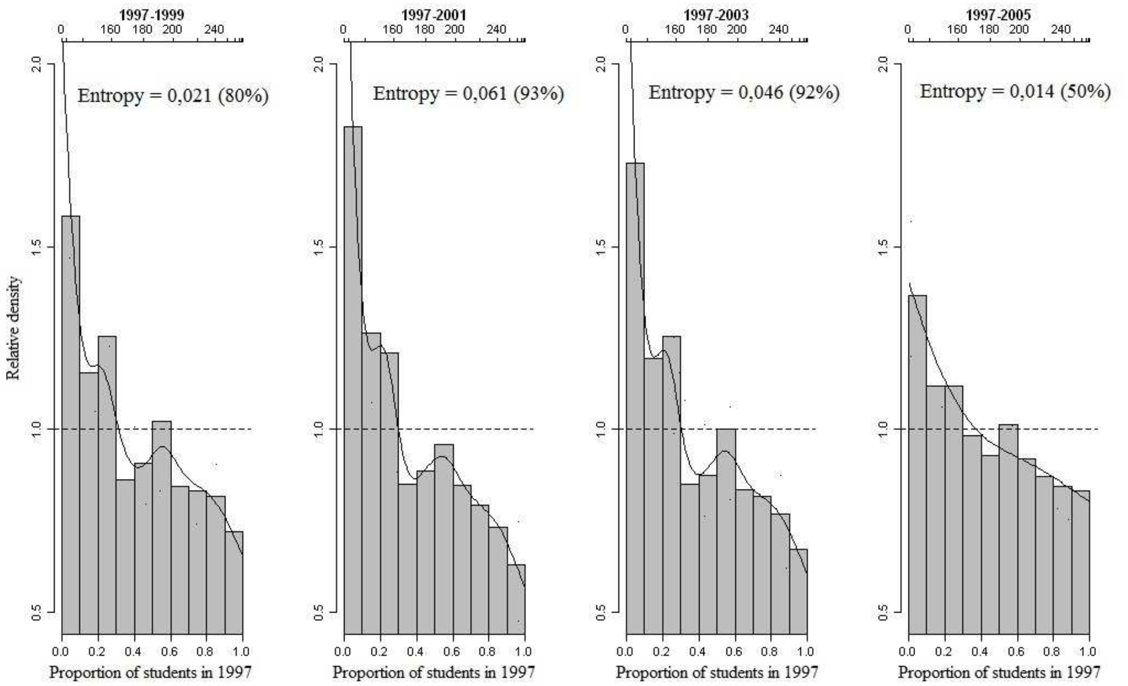
We can perceive in Panel B that the median downshift in test scores was clearly the dominant factor in all periods, as expected. The importance of location effect to explain the dissimilarities between two distributions can be seen not only by the fact that the curve in Panel B practically reproduces the curve in Panel A, but also because of the result of the entropy index indicated in the upper part of the graphs. In decomposing the total entropy index, we can see that the one that originates from location effect represents 50 to 93 percent of the total dissimilarity between the curves, depending on the period analyzed.

Figure 1
Relative distribution and entropy index, fourth grade of elementary school, Math, Brazil, 1997 to 2005

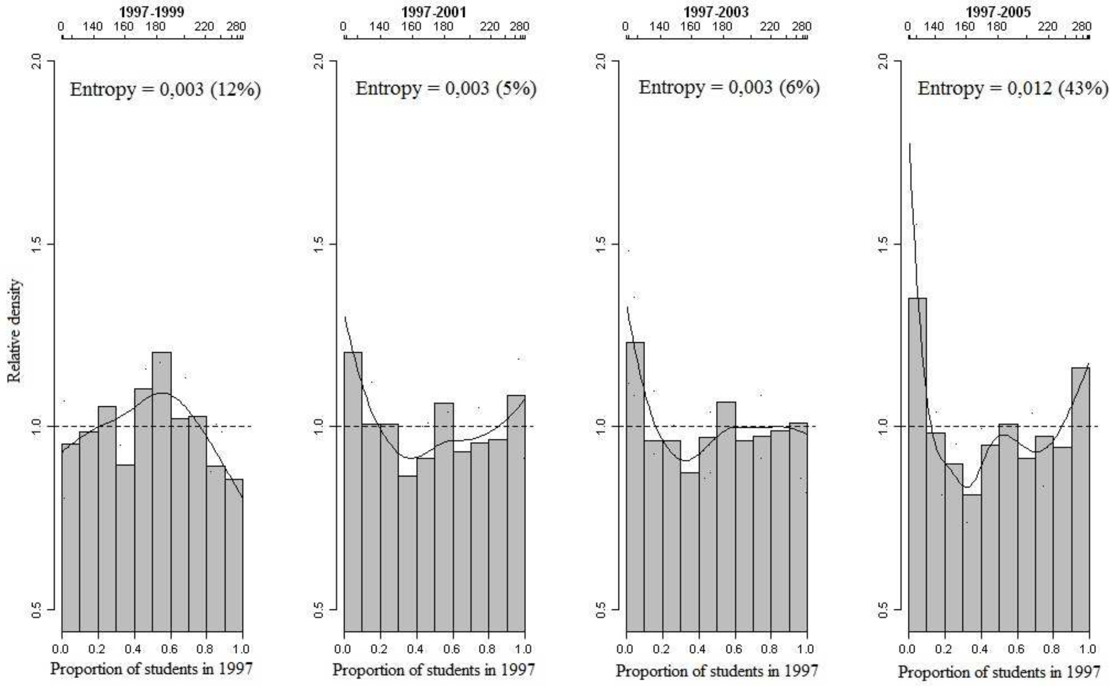
Panel A: Overall Relative Distribution



Panel B: Location Shift



Panel C: Shape Shift



Source: INEP, National System for Evaluation of Basic Education (SAEB), 1997, 1999, 2001, 2003 and 2005.

These results suggest that at the end of the 1990's and the beginning of the twenty-first century a scenario emerged. This scenario suggest a less favorable position was conferred to the students who composed the cohort of students enrolled in the fourth grade of elementary school in the years of comparison, who appeared to have test scores inferior to those found in 1997. In this sense, the average decline in elementary school performance after 1997 seems to be explained, in large part, by a generalized worsening of test scores for all the students enrolled in the fourth grade of elementary school in Brazil.

While the median shift had been the dominant factor, there was also a small polarization trend that was not evident in the overall relative distributions (Panel A), but can be seen clearly in Panel C. Besides the graphs, we can analyze the differences in the shape of the distributions through the polarization indices presented in Table 3. As mentioned, the shape effect is calculated by the comparison between the empirical distribution of the reference year (1997) adjusted to have the same average and empirical distribution of the comparison year. This captures, therefore, the residual changes between the two distributions.

The results show that there are two well-defined behaviors for the polarization. The first of which can be seen between 1997 and 1999. We can perceive a negative polarization that is characterized by the inverted U format in the bars and the smoothened lines of the relative densities. This means that in 1999 the performance of the cohort of students in the fourth grade of elementary school showed a trend of equalization, due to the reduction in the density of students with a much lower performance (inferior tail) and

much higher performance (superior tail) and the rise in the density of students with an average performance.

The magnitude and sign of the polarization index confirm the impression suggested by the graph analysis. The relative polarization index of the median was significant to one percent, which indicates an inverse polarization, i.e., a convergence of the school results to the median of the distribution. In this process, it is possible to perceive that the superior tail had a larger contribution. The polarization index of the superior tail is negative and statistically significant to one percent, while the polarization index of the inferior tail is negative with a low statistical significance.

This result suggests that, between 1997 and 1999, part of the decline in average elementary school performance is explained by residual changes in the superior tail of the distribution that are associated a priori in the educational results of the potentially more able students. This means that, in 1999, the students located in the upper tenths of the distribution had elementary school performances lower than that of students in this same position of the distribution in 1997.

The other result found in the component related to the shape effect can be seen in other periods (1997-2001; 1997-2003; 1997-2005) and is characterized by the existence of a positive polarization, mainly in the inferior tail of the distribution. We can see that the relative polarization index of the median was positive and significant to one percent. The polarization of the inferior tail contributed greatly to this result, since the indices are high and considerably significant. These findings make it evident that after 1997 there is a greater concentration of students with lower elementary school performance, a fact which contributed, in a residual manner, both to the lower average found during the years subsequent to 1997 and to the increase in educational inequality.

Table 3
Polarization index, fourth grade of elementary school, Math, Brazil, 1997 to 2005

Polarization index	Estimates			
	1997-1999	1997-2001	1997-2003	1997-2005
Median (MRP)	-0,043***	0,037***	0,025***	0,057***
Lower tail (LRP)	-0,018*	0,067***	0,056***	0,089***
Upper tail (URP)	-0,067***	0,006	-0,005	0,024***

Source: INEP, National System for Evaluation of Basic Education (SAEB), 1997, 1999, 2001, 2003 e 2005.
Note: *** significant at 1%; * significant at 10%; all others are not significant.

6. Final remarks

Typically, studies that seek to analyze the changes in elementary school performance of students over the years focus on synthesis-measures, such as the mean and median of test scores. Even though these measures are useful in illustrating the trend in the quality

of education, they are incomplete in that they hide various aspects inherent to the distribution. For example, an increase in the average elementary school performance should be interpreted as a positive change in the quality of education if it originates from an overall improvement in the performance of all students and not only from a specific group of students, like those who are already well placed in the elementary school performance scale. In this last case, the rise in average quality would be followed by a rise in the inequality of elementary school results.

The purpose of this paper was to explore the changes behind the distribution of elementary school performance of Brazilian students that resulted in at decline of averages after 1997. To do this, we used a non-parametric methodology called relative distribution, developed by Handcock and Morris (1999) in order to study income inequalities in the job market, and we applied it to educational data. We used data from SAEB for the period from 1997 to 2005, for a cohort of students enrolled in the fourth grade of elementary school and evaluated on math tests.

The decline in average elementary school performance at the end of the 1990’s and the beginning of the twenty-first century in Brazil can be attributed to two sources: 1) an overall fall in the scores of all students, including the scores of those with greater cognitive ability; 2) a fall in the scores of a specific group of students, for example, of those with a lesser cognitive capacity (located in the inferior tail of the distribution) or of those with greater cognitive capacity (located in the superior tail of the distribution).

The results presented in this study show that the decline in average elementary school performance was as a consequence, primarily, of an overall fall in the scores. This means that after 1997, mostly in 1999 and 2001, the performance attained by students of lower capacity, average capacity and those of greater capacity fell. Since the decline in the average was, for the most part, a result of an overall negative performance by all students, it is logical to suppose that the reduction in the average quality of education has been due to a structural problem, that may be associated to the mass influx of students in the second half of the 90’s.

The substantive rise in enrollments occurred mainly in municipal school districts, a phenomenon known as the municipalization of education (Soares and Souza, 2003), and it helped give rise to the Elementary and Secondary School Maintenance and Development Fund (FUNDEF), in 1997. Within all the school districts in Brazil, the public municipal schools, on average, are the ones that most exemplify the precarious nature of the school environment; for example, low teacher qualifications, a poor pedagogical support infrastructure, a high proportion of students with a low social background and, consequently, with greater learning difficulties, amongst others. Within this context, the lack of preparedness to receive a larger contingent of students brought about by policies to expand access to education may have accentuated the problems already faced by these schools. This is reflected in the elementary school results of these students on the SAEB proficiency exams.

The rise of enrollments was propelled by the inclusion of children from less favored social backgrounds. Therefore, it is possible to imagine an increase in the socioeconomic and age group heterogeneity of the classes that may have negatively affected the performance of the students as a whole. In educational literature, this phenomenon is known as peer effect. The underlying idea is that if a greater number of students of less favored socioeconomic conditions who are behind in their grade are introduced into the learning environment, the classroom tends to be a less stimulating learning environment and the students become complacent and their academic performance consequently worsens.

A result in favor of this hypothesis was found in the study by Pinto (2010). In developing a semi parametric methodology to estimate the function of educational production relating to peer effect, the author found evidence of the existence of this effect for all the students, irrespective of their qualities. Pinto (2010) used 2003 data from SAEB, for students tested in mathematics in the fourth grade of elementary school. Another example found for Brazil is relayed in a study by Machado et al. (2008). Using 2003 data from SAEB, for students tested in mathematics and Portuguese language in the fourth grade of elementary school, the authors showed that there is an inverse association between age group dispersion of the classes and individual proficiency. This means that children placed in more heterogeneous classes vis-a-vis age obtain lower results on proficiency exams when compared with children of the same school who are placed in classes that are more homogeneous in age.

Apart from the generalized decline in students' performance, the results of the relative distribution showed that changes in the shape of the distribution, characterized by an increase in the density of the inferior tails, contributed, though residually, to the decline of average performance, mostly, in the year 2005. With the exception of 1999, the shape effect signals an increase in elementary school performance inequality, since there is a growth in the number of students with a low performance (located in the inferior tail of the distribution) concomitant with the rise in the number of students with a higher performance (located in the superior tail of the distribution).

In conclusion, this paper attempted to demonstrate that we can find important aspects behind a test scores distribution. The relative distribution method provides additional insight about the decline of average cognitive skills and it helps us to formulate some possible explanation for this negative trend. As we suggested, the average decline could be associated to direct and indirect effects of changes in socioeconomic status of students due to the increase of enrollments at the end of the 1990's. However, our analysis did not employ controls for socioeconomic status. We have focused on comparing the distributions of a single variable between years. If the socioeconomic status of a cohort of fourth graders varies systematically over time, the impact of this covariate is of interest. The next step will be to extend this analysis to include some important covariate, like socioeconomic status, and set up a better explanation for the relationship between expansion of education and test scores in Brazil.

Acknowledgments

This is a study developed within the scope of the Observatory Education Project and the authors would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for their financial support.

References

ALVES, F. Qualidade da educação fundamental: integrando desempenho e fluxo escolar. **Ensaio: avaliação de políticas públicas educacionais.** Rio de Janeiro, v.15, n.57, p.525-542, out/dez., 2007.

AMMERMÜLLER, A. PISA: **What makes the difference?** Explaining the gap in PISA test scores between Finland and Germany. Centre for European Economic Research, 2004. 29 p. (Discussion Paper, 04-04). Disponível em http://papers.ssrn.com/sol3/papers.cfm?abstract_id=555954. Acesso em 09/01/2008.

BEHRMAN, J. R.; KHAN, S.; ROSS, D.; SABOT, R. School quality and cognitive achievement production: a case study for rural Pakistan. **Economics of Educational Review**, vol. 16, Issue 2, April 1997, pages 127-142.

CARNOY, M. Cuba's Academic Advantage: Why Students in Cuba Do Better in School?

CORAK, M. and LAUZON, D. Provincial differences in high school achievement: for whom do schools matter? Proceedings of Statistics Canada Symposium. Modeling Survey Data for Social and Economic Resarch, 2002.

DING, W. and Leher, S.F. Class size and student achievement: experimental estimates of who benefits and who loses from reductions. 2008.

FERNANDES, R.; NATENZON, P. E. A Evolução recente do rendimento escolar das crianças brasileiras: uma reavaliação dos dados do SAEB. **Estudos em Avaliação Educacional**, São Paulo, n. 28, p. 3-22, dez. 2003.

FERTIG, M. Who's to Blame? The Determinants of German Student's Achievement in the PISA 2000 study. IZA Discussion Paper No. 739, March, 2003. Bonn: Institute for the Study of Labor.

HALL, P.; TAO, T. **Relative efficiencies of kernel and local likelihood density estimators**. London: Blackwell, 2002.

HANDCOCK, M.S.; ALDRICH, E.M. **Applying relative distribution methods in R.** Seattle: Center for Statistics and the Social Sciences, University of Washington, 2002. (Working paper, 27). Disponível em: <http://www.csss.washington.edu/Papers/wp27.pdf>. Acesso em 12 abr. 2007.

HANDCOCK, M.S.; JANSSEN, P.L. Statistical inference for the relative density. **Sociological Methods Research.** Beverly Hills, v.30, n.3, p.394-424, Feb. 2002.

HANDCOCK, M.S.; MORRIS, M. **Relative distribution methods in the social sciences.** New York: Springer-Verlag, 1999.

HANUSHEK, E. Why quality matters in education? In.: Finance and Development. Vol. 42, No 2, June, 2005.

LEVIN, J. For whom the reductions count: a quartile regression analysis of class size and peer effects on scholastic achievement. **Empirical Economics**, 26, 221:246, 2001.

LOADER, C. R. Local likelihood density estimation. **The Annals of Statistics**, Pennsylvania, v. 24, n. 4, p. 1602-1618, Apr.1996.

NERI, M.; CARVALHO, A. **Seletividade e medidas de qualidade da educação brasileira 1995-2001**. Rio de Janeiro: EPGE; FGV, 2002. (Ensaios Econômicos, 463).

N'TCHOUGAN-SONOU, C. Automatic promotion or large-scale repetition: which path in quality? **International Journal of Educational Development**, Vol. 2, Issue 2, January, 2001, p. 149-162.

PINTO, Cristine C. X. Estimation of peer effects in classrooms: evidence for brazilian schools in 2003. Belo Horizonte: Cedeplar/UFMG, 2007. Disponível em: <http://virtualbib.fgv.br/ocs/index.php/sbe/EBE08/paper/viewFile/379/19>. Acesso em: 04 abr. 2009.

SOARES, J. F. . Measuring cognitive achievement gaps and inequalities. **International Journal of Educational Research**, Cambridge v. 45, n. 3, p. 176-187. 2006.

SOARES, S. SOUZA, A. **A demografia das escolas:** decomposição da municipalização do ensino básico. Rio de Janeiro: IPEA, 2003. (Texto para Discussão, 940).

SOUZA, P. R. **Avaliação a serviço da qualidade educativa.** São Paulo: Reescrevendo a Educação, 2006. Disponível em: <www.reescrevendoaeducacao.com.br/2006/pages.php?recid=44-44k>. Acesso em: 05 jul. 2008.

WANG, D. B. (2004). Family background factors and mathematics success: A comparison of Chinese and US students. **International Journal of Educational Research**, *41*, 40-54.

WOESSMANN, L. How Equal Are Educational Opportunities? Family Background and Student Achievement in Europe and The US. CESinfo Working Paper No. 1162,Catgegory 4: Labor Markets, March, 2004.

WORKING PAPER ~~8~~/2011 • MAY, 2011 •