The Decline of Test Scores in Brazil between 1995 and 2003: Socioeconomic, Private/Public Schools and Residual Effects

The Decline of Test Scores in Brazil betweem 1995 and 2003: Socioeconomic, Private/Public Schools and Residual Effects

Clarissa Guimarães Rodrigues[♣]
Eduardo Luiz Gonçalves Rios-Neto[♦]

SHORT ABSTRACT

This paper deals the decline of test scores between 1995 and 2003 through several types of decompositions. It starts with an application of a decomposition suggested by Handock and Morris (1999) focused solely on the scores (dependent variable). The main goal of this exercise is to see if the observed decline is due to a change in the shape of score distribution or due to a general decline in proficiency. It moves to a second decomposition applying the Oaxaca-Blinder decomposition framework, in this case the predicted score is decomposed in the effect due the vector of socioeconomic variables and the effect due to the vector of betas (the "price" effect of the school system valuating the socioeconomic variables into proficiency scores). A third decomposition incorporates the residuals using the approach developed by Juhn, Murphy, and Pierce (1993). Finally, a JMP (1991) decomposition is performed applying differences in differences, so that students in private and public schools are in different groups. Results based on Reading Scores at the eighth grade indicate that the observed decline in proficiency is not due to a compositional effect associated with school expansion in Brazil, when students from low SES are incorporated in the system. Rather, the score differentials in time are associated with the role of betas wich tend to be linked on how schools transform person attributes into performances. When the differentials are separated between private and public schools, then the residual component becomes important explaining the increasing performance gap between these two types of schools, a residual component that is associated with unobserved students´ personal attributes.

EXTENDED ABSTRACT

This paper deals the decline of test scores between 1995 and 2003 through several types of decompositions. It starts with an application of a decomposition suggested by Handock and Morris (1999) focused solely on the scores (dependent variable). The main goal of this exercise is to see if the observed decline is due to a change in the shape of score distribution or due to a general decline in proficiency. It moves to a second decomposition applying the Oaxaca-Blinder decomposition framework, in this case the predicted score is decomposed in the effect due the vector of socioeconomic variables and the effect due to the vector of betas (the "price" effect of the school system valuating the socioeconomic

♣ Graduate Student of Demography at Cedeplar/UFMG.

♦ Full Professor of Demography at Cedeplar/UFMG.

variables into proficiency scores). A third decomposition incorporates a decomposition of the residuals using the approach developed by Juhn, Murphy, and Pierce (1993).

1- Handcock and Morris Decomposition

The first decomposition follows the following model:

Y_0 = test score results in year t=0 (reference group)
Y_1 = test score results in year t=1 (comparison group)

Denote the cumulative distribution function (CDF) of:
$Y_0 \rightarrow F_0(y)$
$Y_1 \rightarrow F_1(y)$

The relative distribution of Y_1 to Y_0 is defined as the distribution of the random variable:

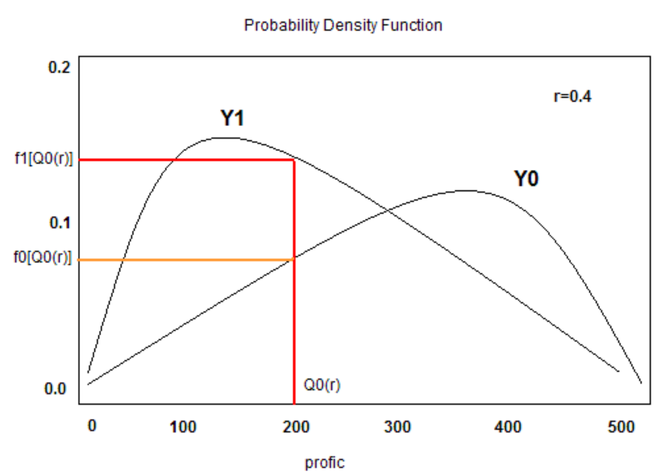
$$R = F_0(Y_1)$$

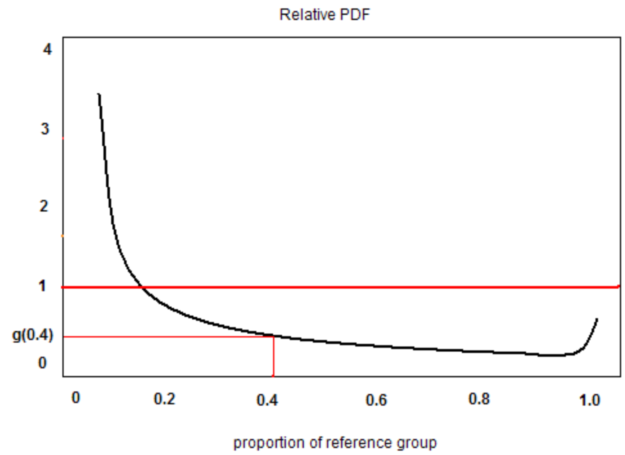
As a random variable, R has both a CDF and a PDF (relative density):

CDF $\rightarrow$ $$G(r) = F_1\left(F_0^{-1}(r)\right) = F_1(Q_0(r)) \qquad 0 \leq r \leq 1$$

$Q_0(r)$ is the quantile function of F_0

PDF (relative density) $\rightarrow$ $$g(r) = \frac{f_1(Q_0(r))}{f_0(Q_0(r))} \qquad 0 \leq r \leq 1$$





Let Y_{0L} denote a random variable describing the reference group *location-adjusted* to have the same mean as the comparison group.

Y_{0L} defines a hypothetical group which has the location of the comparison group, but the shape of the reference group.

For an additive mean shift, we define Y_{0L} as a random variable $Y_{0L} + \rho$ where $\rho = \mu Y_1 - \mu Y_0$

CDF of $Y_{0L} \rightarrow F_{0L}(y) = F_0(y - \rho)$

From these three distributions – Y_0, Y_{0L}, Y – we can construct two Relative Distributions that represent the effects of the location and shape changes.

$R_0^1 = F_0(Y_1)$ = relative distribution of Y_1 compared to Y_0;

$R_0^h = F_0(Y_h) = F_0(Y_0 + p)$ = relative distribution of Y_h compared to Y_0 defining the pure level effect;

$R_h^1 = F_h(Y_1)$ = relative distribution of Y_1 compared to Y_h defining the structure effect.

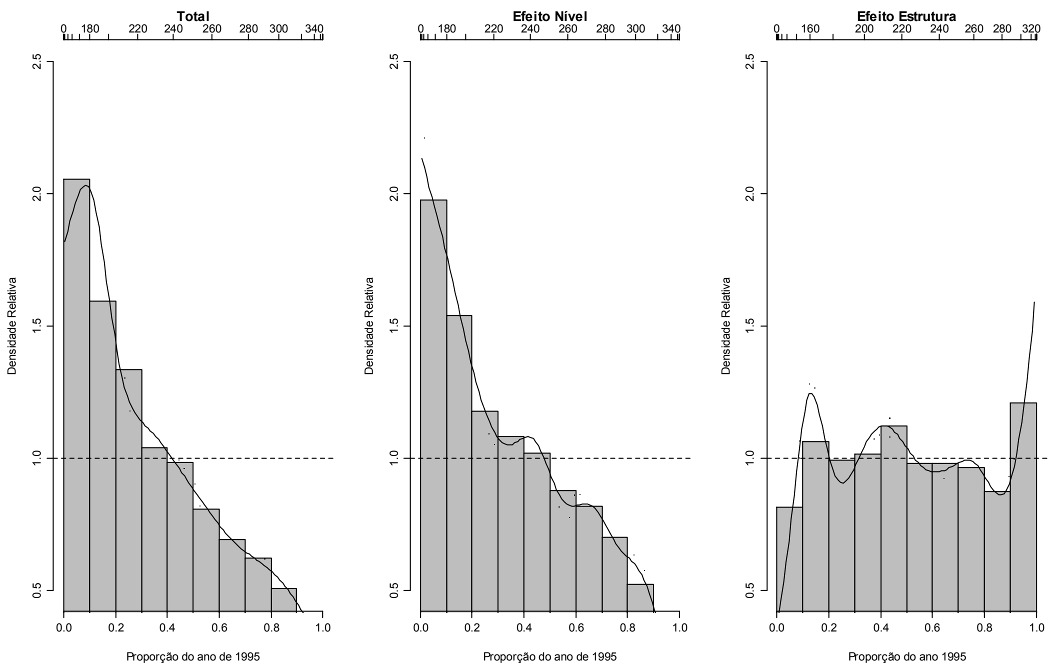
R_0^h will present a uniform distribution when reference and comparison groups have the same level. R_h^1 has a uniform distribution when controlling for the level effect both reference and comparison groups will have the same structure.

This can be represented in terms of the density ratios:

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

The figure 1 below shows an example of this first decomposition in the dependendent variable applied to

Figure 1 - Relative Distribution of Reading Scores and Decomposition, Total, Level, and Shape Effects, Eighth Grade, Brazil, 1995 e 2003



Source: National System for Evaluation of Basic Education in Brazil (SAEB), 1995 e 2003.

The overall decline of 24.1 points in the reading score at eighth grade is primarily due to a level effect (compare the middle with the left graph in Figure 1), and the change in shape (right graph in Figure 1 compared with left graph) is not very pronounced. This is an important result once there was an expansion in Brazilian school coverage during this period, this expansion was comprised by the inclusion of low SES status students in the system. Specialists in the field of education argued that the decline in the overall score during this period was due to a mere compositional (shape) effect, a point not confirmed by this exercise.

2- Oaxaca-Blinder Decomposition

One way to explore the scores differential between groups is to decompose it into "explained" and "unexplained" components.

Assume that proficiency level in math for individual i in group 1 can be written as:

$$P_{1i} = \beta_1 X_{1i} + \mu_{1i} \tag{1}$$

and proficiency level in math for individual i in group 2 can be written as:

$$P_{2i} = \beta_2 X_{2i} + \mu_{2i} \tag{2}$$

The difference in mean proficiency can be written as:

$$P_1 - P_2 = (X_1 - X_2)\beta_1 + (\beta_1 - \beta_2)X_2 \tag{3}$$

The first term in this decomposition represents the “explained” component, that due to average differences in background characteristics of pupil from groups 1 and 2. The second term is the “unexplained” (not to confuse with the non-explained component that is the residual in a regression) component, and represents differences in the estimated coefficients, i.e., differences in the returns or school perfomance due to similar characteristics between groups 1 and 2.

A decomposition exercise of the reading scores at eighth grade is shown in Table 1. The exercise controls for SES variables such as mothers´ education, gender, and race/color, familiar structure, as well as for public/private schools. The total decline in predicted reading score during the period is aroound 18.15 points. Almost all predicted change in the score is due to the “unexplained” component which is actually a beta effect. The implication is that the SES attributes that could be related with the expansion in school coverage are indeed not important for the explanation of grade proficiency in Brazil. It seems that some other factors associated with the way the schools operate in the country might be important factors behind this explanation. Indeed, when one looks at the change in beta associated with public/private schools in Table 1, the shift in this component is the most important determinant of this decline.

Table 1 -Oaxaca-Blinder Decomposition for Reading Scores, Eight Grade, Brazil, 1997 and 2003

Variables	Oaxaca-Blinder decomposition results	
	Quantity effect	Price effect
sex (0 = female; 1 = male)	-0,163	-4,043
race (0 = white; 1 = black)	-0,409	-1,468
age grade gap (0 = no; 1 = yes)	-0,141	3,753
type of school (0 = particular; 1 =public)	-0,944	-6,310
Familiar Structure (*live with both parents = omitted*)		
only mother	0,110	-0,386
another relative	-0,129	-0,800
Mother schooling (*0 years of schooling = omitted*)		
1 to 4 years of schooling	0,326	0,226
5 to 8 yers of schooling	-0,458	0,118
9 to 11 years of schooling	-0,100	0,692
11+ years of schooling	-0,979	0,634
Unknown schooling	0,111	0,139
Region (*Northeast = omitted*)		
North	0,011	0,144
Southest	-0,326	-0,068
South	-0,238	-0,247
Centre-West	0,019	-0,237
Constante	-	-6,987
Total	-3,310	-14,841
Mean prediction 1997	252,222	
Mean prediction 2003	234,072	
Y_{03} - Y_{95} predicted	-18,150	

Source: National System for Evaluation of Basic Education in Brazil (SAEB), 1997 e 2003.

3- Juhn, Murphy, and Pierce (1993) Decomposition

The Oaxaca-Blinder decomposition calculated above deals only with predicted score, ignoring what is happening with the residuals. A useful framework for isolating the observable and unobservable effects is to write a simple wage equation such as:

$$Y_{it} = X_{it}\beta_t + u_{it}$$

It will be useful to think of this residual as two components: an individual's percentile in the residual distribution, θ_{it}, and the distribution function of the proficiency equation residuals, $F_t(\ )$. By definition of the cumulative distribution function:

$$u_{it} = F_t^{-1}(\theta_{it} / X_{it})$$

where $F_t^{-1}(./X_{it})$ is the inverse cumulative residual distribution for pupils with characteristics X_{it} in year t.

In this framework changes in overall scores´ inequality come from three sources:

1) changes in the distribution of individual characteristics;
2) changes in the prices of observable skills;
3) changes in the distribution of residuals.

The decomposition is illustrated by the formula:

$$Y_{it} = X_{it}\overline{\beta} + X_{it}(\beta_t - \overline{\beta}) + \overline{F}^{-1}(\theta_{it} / X_{it}) + [F_t^{-1}(\theta_{it} / X_{it}) - \overline{F}^{-1}(\theta_{it} / X_{it})]$$

The first term captures the effect of a changing SES attributes and other covariates such as public/private schools evaluated at fixed betas (prices or performance coefficients). The second term captures the effects of changing performance coefficients for observables at fixed X's, and the final term captures the effects of changes in the distribution of proficiency residuals.

The proficiency distribution can be reconstructed under this framework:

- fixed betas and residuals: $Y_{it}^1 = X_{it}\overline{\beta} + \overline{F}^{-1}(\theta_{it} / X_{it})$
- fixed residuals: $Y_{it}^2 = X_{it}\beta_t + \overline{F}^{-1}(\theta_{it} / X\theta_{it})$
- all components changes through time:

$$Y_{it}^3 = X_{it}\beta_t + F_t^{-1}(\theta_{it} / X_{it}) = X_{it}\beta_t + u_{it} = Y_{it}$$

The basic decomposition technique will be to calculate the distribution of Y_{it}^1, Y_{it}^2 e Y_{it}^3 for each year and attribute the change in inequality through time in the Y_{it}^1 distribution to changes in observable quantities. We then attribute any additional change in inequality in Y_{it}^2 to changes in observable prices (betas), finally we attribute any additional changes in inequality for Y_{it}^3 beyond those found for Y_{it}^2 to changes in the distribution of unobservable (i.e., changes in unmeasured prices and quantities).

Table 2 shows that the decompostion evaluated at the mean value is not very different from the results obtained in the Oaxaca-Blinder decomposition, with a minor residual component in the decomposition. Nevertheless, the residual component is relatively more important when the decomposition is performed at different quantile levels of the distribution.

Table 2 - Juhn, Murphy and Pierce Decomposition for Reading Scores, Eight Grade, Brazil, 1997 and 2003

Mean and Percentiles	Total difference	Quantities effect	Prices effect	Unobservable quantities and price effects
Mean	-18,151	-3,292	-14,841	-0,018
p5	-14,273	-0,992	-15,889	2,608
p10	-18,225	-2,853	-15,541	0,169
p25	-19,621	-2,707	-15,674	-1,240
p50	-20,448	-4,246	-15,465	-0,736
p75	-18,042	-3,856	-14,070	-0,117
p90	-14,494	-1,430	-13,669	0,606
p95	-14,820	-3,115	-13,017	1,312

Source: National System for Evaluation of Basic Education in Brazil (SAEB), 1997 e 2003.

4 Juhn, Murphy, and Pierce (1991) Decomposition

Another way to expand the Oaxaca-Blinder decomposition in order to incorporate the residual term is to follow Juhn, Murphy, and Pierce (1991), working in termos of a difference in differences decomposition. In this case we will specify a model separating public and private schools, since these two dimensions tend to be completely different and its dummy variable was very important in the previous decomposition.

$$D_t = P_{Ait} - P_{Bit} = \underbrace{(X_{Ait} - X_{Bit})\beta_{At}}_{\text{Explained}} + \underbrace{(\theta_{Ait} - \theta_{Bit})\sigma_t}_{\text{Non-Explained}}$$

P_{Ait} = score for private schools at year t,

P_{Bit} = score for public schools at year t.

JMP (1991) decomposition between t and t+1:

$$D_{t+1} - D_t = [(X_{Ait+1} - X_{Bit+1}) - (X_{Ait} - X_{Bit})]\beta_{At} + (X_{Ait+1} - X_{Bit+1})(\beta_{At+1} - \beta_{At}) +$$
$$[(\theta_{Ait+1} - \theta_{Bit+1}) - (\theta_{Ait} - \theta_{Bit})]\sigma_t + (\theta_{Ait+1} - \theta_{Bit+1})(\sigma_{t+1} - \sigma_t)$$

$$D_{t+1} - D_t = (\Delta X_{t+1} - \Delta X_t)\beta_{At} + \Delta X_{t+1}(\beta_{At+1} - \beta_{At}) +$$
$$(\Delta\theta_{t+1} - \Delta\theta_t)\sigma_t + \Delta\theta_{t+1}(\sigma_{t+1} - \sigma_t)$$

- first component: observed quantity effect
- second component: observed price or beta effect
- third and fourth comopnents: residual effect decomposed into quantity (unobserved) effect and price (beta) unobserved effect.

σ_t = standard deviation at year t,

σ_{t+1} = standard deviation at year t+1.

θ_{Ait} = standardized residual at private school in year t;

θ_{Bit} = standardized residual at public school in year t;

θ_{Ait+1} = standardized residual at private school in year t+1;

θ_{Bit+1} = standardized residual at public school in year t+1;

The implied assumption by the authoris is that:

- $\sigma_{At} = \sigma_{Bt} = \sigma_t$
- $\sigma_{At+1} = \sigma_{Bt+1} = \sigma_{t+1}$

The decomposition presented in Table 3 is for a different period than the previous one (1997-2003), but it will be recalculated to be fully compatible. The decomposition exercise shows that the change in the difference between the scores of students in private and public schools in the two periods is positive (9,4 points). This means that the public school performance is relatively declining in through time. The positive predicted gap went the same direction increasing the private and public school gap in 1.69 points, most of this decline would be explained by the quantity effects. The residual effect contributed to increase the private and public school in 7.76 points. In this case, the most important component of the residual gap is associated with the increasing difference in the vector of socioeconomic variables characteristics of students in private versus public schools.

Table 3 - Juhn, Murphy and Pierce Decomposition for Reading Scores, Eight Grade, Brazil, 1997 and 2003

Decomposition of individual differentials	Raw differential	Quantity effect	Residual gap	
1997	42,100	17,671	24,429	
2003	51,532	19,340	32,192	
Difference in components of differentials	Diference in differential	Diference in predicted gap	Difference in residual gap	
Total	9,433	1,669	7,764	
Decomposition of difference in predicted gap	Difference in predicted gap	Quantity effect	Price effect	Interaction quantity vs. price effect
Total	1,669	0,963	0,126	0,580
Decomposition of difference in residual gap	Difference in residual gap	Quantity effect	Price effect	Interaction quantity vs. price effect
Total	7,764	7,168	0,554	0,040

Source: National System for Evaluation of Basic Education in Brazil (SAEB), 1997 e 2003.

References

BLINDER, Alan (1973). "Wage Discrimination: reduced forms and structural estimation". Journal of Human Resources, 8, 436-455.

JUHN, Chinhui; MURPHY, Kevin M.; and PIERCE (1991) Brooks. "Accounting for the Slowdown in Black-White Wage Convergence." In *Workers and Their Wages,* edited by M.H. Kosters. Washington D.C.: American Enterprise Inst. Press.

JUHN, Chinhui, Kevin MURPHY and Brooks PIERCE (1993). "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, vol. 101, pp.410-42.

HANDCOCK, M.S. e MORRIS, M. **Relative distribution methods in the social sciences.** Springer-Verlag: New York, 1999. Caps. 1 a 8.

OAXACA, Ronald (1973). "Male-Female Wage Differentials in Urban Labor Markets". International Economic Review, 14, 693-709.

.