Social and school determinants of educational achievement in Brazil, 2007

Raquel Guimaraes[1]

José Francisco Soares[2]

Raquel Alvares[3]

Izabel Fonseca[4]

Abstract

This study aims to assess the determinants of 4th and 8th grade public school students' educational achievement in Brazil. We employ hierarchical linear regression models to evaluate the effect of personal attributes, socioeconomic factors and school and teachers characteristics on the Reading and Math proficiency. We use microdata from Prova Brasil 2007, which evaluates the performance of all Brazilian public schools' students in 4th and 8th grade. Our evidence shows that family characteristics and peer effects, expressed by the school socioeconomic level and the proportion of lagged students, had presented a significant impact on the student's proficiency. Finally, the characteristics of schools, teachers and principals have had also a positive effect on the educational achievement, but this impact on students" score is small compared to personal, family and peer effects.

Extended Abstract

This paper proposes the use of summary measures of socialeconomic background and school characteristics in order to assess the determinants of educational achievement of 4th and 8th graders in 2007. The work with summary measures has the advantage of capturing a latent trait in the population being studied by several indicators in the Prova Brasil questionnaire. Alves and Soares (2009) present in detail the theoretical motivation for the construction of summary measures.

In this work, we used a model of Item Response Theory - IRT (Hambleton, 1993) in order to aggregate the information in the questionnaire available in Brazil and synthesize the information into a single latent trait. The IRT includes a series of models whose main objective is to obtain measures of latent constructs based on dichotomous and/or ordinal variables. Specifically, we used the model of Samejima (1969), suitable for items with graded or ordinal responses.

[1] Stanford University and CEDEPLAR/UFMG - raquelg1@stanford.edu

[2] School of Education, Federal University of Minas Gerais - francisco-soares@ufmg.br

[3] Group of Evaluation and Educational Measures, Federal University of Minas Gerais - raquelpes@gmail.com

[4] Group of Evaluation and Educational Measures, Federal University of Minas Gerais - izabelfonseca@gmail.com

It should be noted that the explanatory power of the variables which compose the latent trait can be validated by statistical procedures. In this paper we estimate the polychoric matrices, analyze their eigenvalues and eigenvectors, and also analyze item characteristic curves and item information.

Regarding the explanatory model, we use hierarchical models. These models have important methodological advantages over traditional ones, mainly because they allow, when working with variables of the student, to avoid the loss of important information that would occur in the school level. At the same time, hierarchical models allow the incorporation of group variables - in this case, the schools, when considering the existence of variability of both the intercept and slopes (Cameron and Trivedi, 2005, Raudenbush and Bryk, 2000).

For this study, we used hierarchical models with fixed effects for the explanatory variables and a random effect on the intercept which varies from school to school, known as the *random intercept model*. Our model presents the variables for the scores of the constructs that measure aspects of the student's family, school quality, teacher and director of each student. These variables were transformed to a scale of 0 to 10 where 0 is the worst score for the construct and the best score is 10.

In hierarchical modeling, it is usual to center the variables in order to identify the precise effect on the outcome variable. In this study, the student's SES is a measure between 0 and 10, where 0 indicates the lowest socioeconomic level and 10 indicates the highest level. To obtain a measure of socioeconomic status of the school, we calculate the average SES of the students per school. We would only assume the present value of 0 in the School_SES variablw if all students in the school have a score of zero for the NSE. At the same time, the School_SES will be 10 if and only if all students from the same school have the best score in the NSE. These two situations are extremely rare. Thus, as the School_SES most likely does not reach the values 0 and 10, then this variable is centered around the overall average SES of all schools, giving rise to variable c_School_SES. This was also the procedure adopted for the variable which indicates the mean student's lag in each school, school_lag, which is centered on the proportion of lagged studens of all schools.

Table 1: Results of the Hierarchical Linear Model. Dependent variable: Prova Brasil score in Reading and Math. 4th and 8th graders, 2007

Independent variables	4th grade		8th grade	
	Reading	Math	Reading	Math
Men	-7.78***	3.37***	-11.57***	9.50***
	(0.059)	(0.063)	(0.069)	(0.068)
White	8.25***	9.08***	7.54***	8.73***
	(0.100)	(0.106)	(0.118)	(0.117)
Brown	8.36***	9.04***	3.57***	4.53***
	(0.096)	(0.102)	(0.113)	(0.111)
Presents school lag	-15.92***	-15.37***	-17.14***	-18.26***
	(0.071)	(0.076)	(0.078)	(0.077)
SES	-0.44***	0.20***	-0.07**	0.47***

	(0.026)	(0.029)	(0.029)	(0.028)
c_SES_school	11.83***	11.90***	9.30***	9.66***
	(0.136)	(0.162)	(0.126)	(0.139)
c_school_lag	2.65***	-0.07	5.25***	1.48***
	(0.442)	(0.506)	(0.508)	(0.559)
Cultural Environment	2.06***	1.89***	3.49***	2.79***
	(0.028)	(0.031)	(0.033)	(0.031)
Parental involvement	3.17***	2.71***	-1.70***	-2.15***
	(0.031)	(0.034)	(0.034)	(0.034)
Student motivation	-0.38***	-0.54***	2.14***	3.51***
	(0.024)	(0.025)	(0.032)	(0.028)
School Environment (Principal)	0.96***	1.32***	0.53***	1.00***
	(0.062)	(0.079)	(0.069)	(0.080)
School Environment (Teachers)	1.09***	1.25***	0.65***	0.64***
	(0.078)	(0.093)	(0.068)	(0.075)
Director Evaluation by Teachers	0.33***	0.46***	0.32***	0.42***
	(0.059)	(0.075)	(0.069)	(0.078)
Quality of the Library	0.29***	0.42***	0.21***	0.28***
	(0.063)	(0.066)	(0.068)	(0.075)
Intra-school Cohesion	0.12*	0.04	0.24***	0.18**
	(0.064)	(0.072)	(0.067)	(0.072)
Operating Conditions (Principal)	1.01***	1.09***	0.49***	0.51***
	(0.073)	(0.078)	(0.072)	(0.083)
Operating Conditions (Teachers)	0.62***	0.84***	0.50***	0.71***
	(0.065)	(0.073)	(0.052)	(0.057)
Equipment	0.15**	0.43***	0.18***	0.42***
	(0.065)	(0.077)	(0.072)	(0.078)
Facilities	0.24***	0.25***	0.10	-0.13*
	(0.060)	(0.070)	(0.068)	(0.075)
Principal's Training	0.41***	0.44***	0.18***	0.22***
	(0.056)	(0.063)	(0.065)	(0.070)
Teacher uses technology (ICT)	0.71***	0.75***	0.33***	0.32***
	(0.042)	(0.046)	(0.047)	(0.042)
Participatory learning resources	0.35***	0.27***	0.42***	-0.13***
	(0.032)	(0.037)	(0.045)	(0.048)
Formal teaching resources	-0.58***	-0.07*	-0.69***	0.29***
	(0.033)	(0.040)	(0.046)	(0.050)
Teacher's Training	0.22***	0.21	0.22***	-0.19***
	(0.034)	(0.036)	(0.045)	(0.046)
Teacher's Working Condition	0.55***	0.72***	0.55***	0.58***
	(0.036)	(0.042)	(0.041)	(0.044)
Constant	116.20***	118.35***	194.94***	185.79***
	(0.887)	(1.006)	(1.002)	(1.077)

Standard deviation of the constant	11.18	13.25	10.97	12.70
	(0.057)	(0.064)	(0.065)	(0.070)
Residual standard deviation	36.42	38.52	39.64	38.36
	(0.021)	(0.022)	(0.024)	(0.024)
Intraclass correlation coefficient	0.086	0.106	0.071	0.099
Number of students	1,569,662	1,565,392	1,396,086	1,359,375
Number of schools	37,263	37,244	27,377	27,378
Log likelihood	-7,896,882.7	-7,966,353.7	-7,138,132.2	-6,909,710.8

Source: Microdata from Prova Brasil 2007 (INEP/MEC)

Observations: Standard errors between parenthesis

* Significative at 10%

** Significative at 5%

*** Significative at 1%

Table 1 shows that the majority of the estimated parameters are statistically significant at a 1% level. The constant of the model corresponds to the average proficiency among all students. With respect to the explanatory variables, we observe from Table 1 that for the case of Reading, the estimated coefficient for the indicator variable Male is negative, indicating that boys had, everything else held constant, lower performance in reading than girls. However, for performance in mathematics, boys showed a slightly better performance than girls. In the case of race/color, evidence shows that white and brown performed better than blacks. The school lag variable shows a negative impact on the proficiency of the student in both reading and mathematics.

In relation to socioeconomic status, Table 1 reveals for every one-unit increase in score for this variable, proficiency increases little. This relatively small effect may be due to the fact that the variability of SES among students is quite low, since the study deals with Brazilian public school students.

Our analysis also reveals that the variables that show the socio-economic composition of school (c_NSE_escola) and the composition of the school in terms of students lagging behind (c_school_lag) are critical to explain student achievement if we consider the magnitude of the coefficients, everything else held constant. This result shows the importance of the effect of peer socioeconomic status on the proficiency of each student.

The question is how easy it would be for a school to raise its average level in a unit in relation to the overall average? Such question can be answered by analyzing the distribution of this variable. To exemplify the effort to change the student's performance, Figure 1 shows the result of kernel density estimation for the socioeconomic status of the school: it can be seen that the major part of density is located within 1 deviation from the mean. Figure 2 shows that the same can be observed for the proportion of students lagging behind in school: the distribution of schools is between -0.5 and 0.5 from the average and therefore a change of one unit would require considerable effort. Thus, although the effect of SES of peers is important - the magnitude of the effect such a change is not easy to be achieved by schools.

Last but not least, the proportion of students lagging behind in school has a negative impact on student performance. Thus, it is clear the importance of the effect of peers on school performance.

Interestingly, the characteristics of the school, teachers and principals have a marginal impact on performance compared to the impact of personal attributes and family: almost all coefficients for the variables of school, teachers and principals are close to zero, although significant . That is, each increase of 1 unit in the score of these variables, the performance increases by less than 1 point in both Reading and Mathematics and for the 4th and 8th grade. This evidence goes in the direction of the findings in the literature. However, the effort to achieve this marginal impact is not as great. Note, for example, Figure 1, which shows the distribution of the score for the quality and availability of equipment at school. It can be seen that the data are well distributed in the range 00 to 10, with a relatively large variability. Thus, it is difficult for a school to achieve improvements in this score. This has great relevance for policy. But we must keep in mind that, although the effort to improve the characteristics is not large, the impact on school performance will still be marginal.

Figure 1

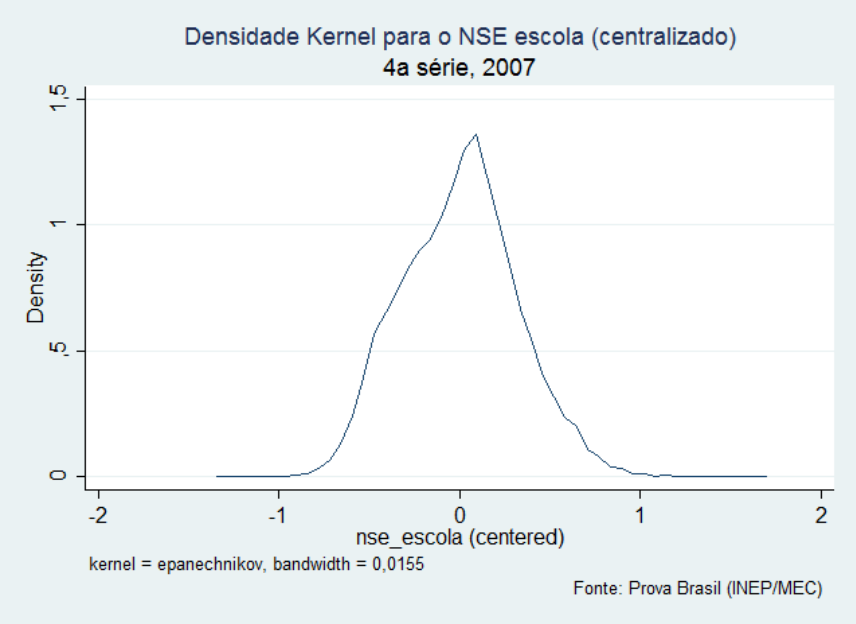


Figure 2

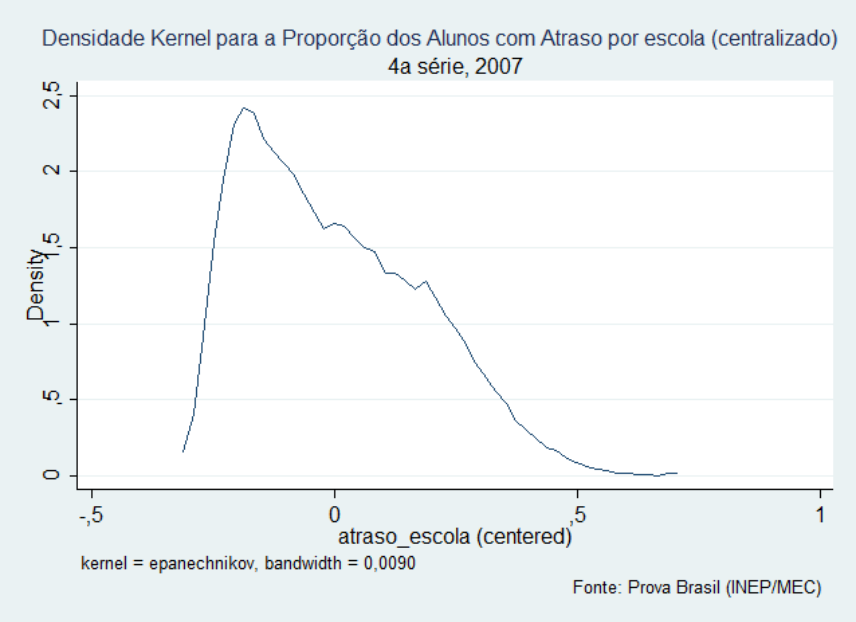
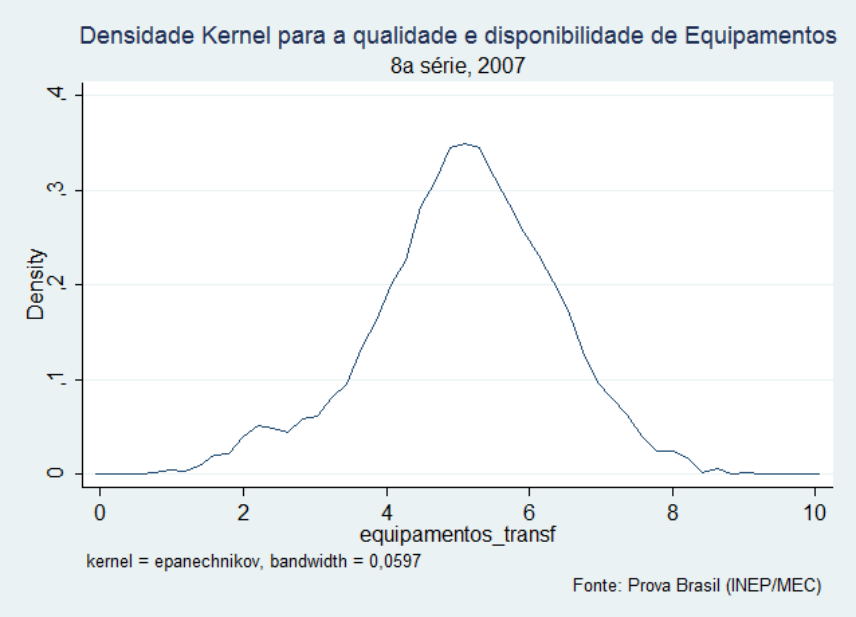


Figure 3



References

ALVES, M. T. G; SOARES, J. F. Medidas de nível socioeconômico em pesquisas sociais: uma aplicação aos dados de uma pesquisa educacional. OPINIÃO PÚBLICA, Campinas, vol. 15, nº 1, Junho,

2009, p.1-30

CAMERON, A.C., TRIVEDI, P.K. 2005. Microeconometrics: Methods and Applications, Cambridge University Press.

HAMBLETON, R. K. Principles and selected applications of Item Response Theory. In: LINN, R. L. (Ed.). Educational measurement. 3. ed. Washington, DC: American Council on Education and the National Council on Measurement in Education, p.147-200, 1993.

RAUDENBUSH, Stephen W; BRYK, Anthony S. Hierarchical linear models: applications and data analysis methods. 2.ed. Thousand Oaks, Calif.; London: Sage, 2000. 485 p.