Volume 50, Issue 4, November 2010 ISSN 1062-9769

ELSEVIER

THE QUARTERLY REVIEW OF ECONOMICS AND FINANCE

(Contents continued on outside back cover)

Published for the Bureau of Economic and Business Research,
University of Illinois at Urbana-Champaign

The Quarterly Review of Economics and Finance 50 (2010) 395–407

Contents lists available at ScienceDirect

The Quarterly Review of Economics and Finance

journal homepage: www.elsevier.com/locate/qref

Barriers to skill acquisition in Brazil: Public and private school students performance in a public university entrance exam[☆]

Tiago Cavalcanti[a,c,*], Juliana Guimaraes[b,c], Breno Sampaio[d]

[a] *Faculty of Economics, University of Cambridge, Sidgwick Avenue CB3 9DD, United Kingdom*
[b] *Newnham College, University of Cambridge, United Kingdom*
[c] *PIMES/UFPE, Brazil*
[d] *Department of Economics, University of Illinois, United States*

ARTICLE INFO

Article history:
Received 7 July 2009
Received in revised form 5 May 2010
Accepted 3 August 2010
Available online 12 August 2010

JEL classification:
J01
J24
O11

Keywords:
Inequality
Public and private schools
Test scores

ABSTRACT

This paper uses a novel data set to quantify the difference in performance of public and private school students in an entrance test exam of the major public university in Brazilian Northeast (Universidade Federal de Pernambuco – UFPE). Although there are many public universities in Brazil, from our knowledge, there is no study that uses data on entrance test scores at such universities to evaluate the determinants of students' performance and the barriers for public school students to get in the good universities. The data set has detailed information on individual and school characteristics, and family background. We found that test scores of public school students are on average about 4.2–17% lower than those taken by private school students, depending on the set of controls. This result is robust when we address problems related to attrition, omitted variables (e.g., cognitive ability), and unobservable selectivity. We also show that once students get into the university, those from public schools perform as well as those from private schools. In addition, the proportion of public school students that gets into the university is roughly the same as the proportion of students doing the entrance exam. However, there is a strong barrier for public school students to get into high competitive majors. The fraction of students from public schools that gets into high competitive majors such as law, medicine, and electronic engineering is almost null. Our findings provide quantitative evidence to the common view that the Brazilian elitist high education system is an important channel for inequality persistence.

"*The new middle classes have more schooling than their parents; some have gone to mushrooming private universities. But they are less educated than the old middle class that benefited from elitist public universities - and that makes moving into the upper class hard.*" The Economist, Aug 16th, 2007. *Adiós to poverty, hola to consumption.*

1. Introduction

Brazil is one of the most unequal countries in the world. Data from the World Development Indicators (World Bank, 2007) show that among 122 countries Brazil is the fourth most unequal. Only Lesotho, Haiti, and Bolivia have an income Gini index that is higher than what is observed in Brazil. In addition, not only the level of inequality is high in the country, but at the same time there is low social intergenerational mobility by international standards (see Bourguignon, Ferreira, & Menéndez, 2003; Ferreira & Veloso, 2006). For instance, Ferreira and Veloso (2006) show that the probability that a black son of a father from the lowest wage quintile will remain in this quintile is about 47%, which is roughly twice than the analogous probability for a white son. This indeed is a strong evidence of low equality of opportunities in the country. However, while the level of inequality is high in Brazil it has been decreasing since 1995 and this decline has been amplified in the last years.[1]

☆ This paper has benefited from financial support of Conselho Nacional de Desenvolvimento Cientfico e Tecnológico (CNPq-Brazil). We thank two anonymous referees, Werner Baer, Pedro Portugal, Donald Pianto, Eduardo Ribeiro, Yony Sampaio, and Bhaskar Vira for comments, and Todd Elder for sharing STATA routines with us. We have also benefited from comments by audiences at Brazilian Econometric Society Annual Meeting, and at the Workshop at University of Cambridge on Policy, Economy and Development in Latin America. We are responsible for any remaining error.

* Corresponding author.
E-mail addresses: tvdvc2@cam.ac.uk (T. Cavalcanti), jgc29@cam.ac.uk (J. Guimaraes), sampaio2@uiuc.edu (B. Sampaio).

[1] Barros, Carvalho, Franco, and Mendonça (2009) show that the Gini coefficient declined from 0.593 in 2001 to 0.552 in 2007; moreover, the ratio of average

1062-9769/$ – see front matter
doi:10.1016/j.qref.2010.08.001

This paper studies the issue of inequality of opportunities in Brazil. There exists a large literature which shows that public education is favorable for growth because it increases the level of human capital and at the same time it tends to produce a more even income distribution.[2] More egalitarian societies are also associated with less social conflicts and individuals have a lower tendency to report themselves happy when inequality is high (e.g., Alesina, DiTella, & MacCulloch, 2004). Therefore, it is important to study the sources of inequalities and to quantify them. According to Psacharopoulos (1999) the average returns to invest in high education is quite high in Brazil (about 20%) when compared to some developed (e.g., United States) and developing (e.g., Chile) countries. For updated values see Psacharopoulos and Patrinos (2004), who also show that the average returns to investment in high education in Brazil is high compared to other developing countries.[3] In a recent article, Menezes-Filho, Fernandes, and Picchetti (2006) show that although human capital has been growing in Brazil it has not contributed much to reduce wage inequality in the country.

In Brazil there are private and public universities as well as private and public schools. There is no tuition at public universities and they in general have a better quality than their private counterpart, which implies that there is a high competition to get in public universities. Students have to take an entrance exam to be accepted in a public university and they must choose their undergraduate major before taking the entrance exam.[4] Investigating the determinants of scores in this exam allows us to evaluate inequality of education among students, in particular, those that did their secondary degree at a public or private school. Although there are many public universities in Brazil,[5] from our knowledge, there is no study that uses data on entrance test scores at universities to evaluate the determinants of students' performance.[6] Is the difference in *ETS* due to family background and school characteristics? Or is it due to unobservable variables, such as ability and effort? Those are important questions since researchers have shown that test scores that evaluate cognitive skills and knowledge are key predictor of future wages, and this effect has being growing in recent years due to skill biased technological change.[7] We use a unique data set on students' entrance test scores (*ETS*) at Universidade Federal de Pernambuco (UFPE), which is the major University in Brazilian Northeast,[8] to evaluate quantitatively differences in scores between public and private school students. The data set allows also to evaluating students in their first academic year at the university. Do students from public school perform worse than private school students in the university? It might be that only high ability public school students get into the university and therefore they might perform better than private school students.

The dataset used in this article is rich with detailed information on individual characteristics (e.g., race, age, and gender), family background (e.g., mother's years of schooling and family income), previous school characteristics (e.g., public or private, if it has lab classes), and the colleges they applied for. We found that test scores of public school students are on average about 4.2–17% lower than those taken by private school students, depending on the set of control variables.[9] This result is robust when we deal with problems such as attrition, omitted variables (e.g., cognitive ability), and unobservable selectivity.

Section 2 describes in detail how we address the problem of omitted variables and unobservable selectivity. Here we briefly discuss our two approaches to deal with such econometrics problem: Firstly, we use the National Exam for High School Students (*ENEM*), as a proxy for students' cognitive ability. The *ENEM* exam is more general that the Entrance Test Exam and emphasizes logical questions. The problem to use the *ENEM* as a proxy for unobservable characteristics is that this exam is not mandatory and in our sample less than half of the students took this exam. In addition, this exam would also capture important differences between private and public schools. Finally, we also use the technique developed by Altonji, Elder, and Taber (2005) to investigate the potential size of any bias on the estimated coefficient of the public school variable due to unobservable selectivity (see more on this on Section 2). This approach is very attractive in non-experimental data and in the absence of valid instruments. It has also been applied in problems similar to ours. Altonji et al. (2005), for instance, applied this technique to investigate the effects of Catholic schools on students' test scores in the United States and Goyal (2009) use the technique to study the gap in students' test scores between private and public schools in India.[10]

Another important result is that family background (e.g., mother's years of schooling and family income) is a key predictor of test scores. Our econometric exercises, therefore, show that not only public school students have a less favorable background (e.g., lower family income and lower parent's education level) and therefore have a lower score in the entrance test exam, but they also perform worse than private school students in this exam given their observed and unobserved individual characteristics. They perform worse even when we control for the *ENEM* exam or school fixed effects (e.g., school's neighborhood), which would capture a high variation in differences in scores between private and public school students. Interestingly, we also show that once students get into the university, those from public schools perform better than (or as well as) those from private schools. Therefore, it does not seem that public school students lack cognitive skills to perform well in education, but they lack specific preparation that is necessary to do well in the entrance exam. Consequently, either improving the quality

income of the 10% richer and the 40% declined 5.2 percentage points over the same period.

[2] See Galor and Zeira (1993). For a recent reference, see Doepke (2004) and Alexopoulos and Cavalcanti (in press).

[3] Resende and Wyllie (2006) provide a survey and some new estimates for the returns of schooling in Brazil.

[4] In Brazil there is not much optional disciplines in undergraduate courses and some are major specific since the first semester.

[5] In the Education Ministry web site in August 2010 there were 58 Federal Universities listed. There are also some important public universities that are State and not Federal ones.

[6] An exception and interesting work is Emilio, Belluzo, and Alves (2004), who use binary model to study the determinants of access to the university of São Paulo. Waltenberg and Vandenberghe (2007) is also an interesting and related study on equality of opportunity in education in Brazil. They use a different data set and study a different question. They found that in order to implement an equality of opportunity policy across pupils of different socio-economic backgrounds, it is necessary to multiply current spending by 6.8 on the lowest achieving pupils.

[7] See Murnane, Willett, and Levy (1995) and Murphy and Peltzman (1997). See Curi and de Menezes-Filho (2005) for evidences on how test scores affect wages in Brazil.

[8] Ferreira and Veloso (2006) show that intergenerational mobility is lower in the Northeast than in the richest Brazilian region (Southeast).

[9] In a cross-country study using the Programme for International Student Assessment (PISA) student-level achievement database, Fuchs and Woessmann (2007) find that students perform better in privately operated schools, but private funding is not decisive. The difference seems to be explained by the difference in organization in the two systems. Using a decomposition approach McEwan (2008) shows that school reforms might have strong impact on students' outcome. For instance, according to him Chile's large-scale school reforms had a strong impact on test scores gap between indigenous and non-indigenous population.

[10] For two recent applications, see Bellows and Miguel (2009), who study the effect of households that experienced intense war violence on the probability of attending community meetings and voting in Sierra Leone, and Kingdon and Teal (2010), who study the effects of teacher unionization on student achievement in India. See also Altonji et al. (2008) who investigate how the effects of Swan-Ganz Catheterization on mortality in intensive care unit (ICU) patients.

of education in public schools by providing a better preparation for the entrance exam or facilitating access for public school students to get into public universities would improve the equality of opportunities in education for students to acquire a university degree.

Finally, we also show that the proportion of public school students that gets into the university is roughly the same as the proportion of students doing the entrance exam. However, there is a strong barrier for public school students to get into high competitive majors. The fraction of students from public schools that gets into high competitive majors such as law, medicine, and electronic engineering is almost null. For instance, in medicine only 6% of the students are from public schools. This number is equal to 7 and 5 for law and electronic engineering, respectively. On the other hand, for less competitive majors the percent of students that came from public schools is much higher. It is, for instance, more than 65% in Music, 64% in geography, 56% in history, and 50% in domestic economics. Data from the 2005 Brazilian National Household Sample Survey (PNAD), show that electrical and civil engineers make about 3 times more than domestic economists and high school teachers.[11] Given wage differentials among occupations, our empirical evaluations suggest that Brazil's public university system is an important source of inequality persistence.

This paper proceeds as follows. Section 2 defines our empirical strategies and discusses some econometric issues. Section 3 describes the dataset. Section 4 presents our empirical estimates. We use both ordinary least squares (*OLS*) procedure and quantile regression. The later is particularly useful to determine whether there are any heterogeneity in the responses of the entrance test scores to different school systems (public or private). In this section, we also evaluate how students perform in the university. Section 5 concludes.

2. Empirical strategy

It is expected that in an university entrance exam the students' score depends on their effort to study for the exam, cognitive ability (e.g., capacity to learn and pay attention to academic activities), quality of the teachers that prepared the student for this exam, resources available (e.g., access to books and news), environment at home and at school, among other variables.[12]

Our goal is to estimate the difference in the quality of the public and private schools using data from entrance test scores in the major university in Brazil's Northeast. We estimate the following equation:

$$\ln(ETS_i) = \beta_0 + \beta_1 * public_i + \beta_2' X_{1i} + \beta_3' X_2' + \varepsilon_i, \quad (1)$$

where ETS_i is the entrance test score of student *i*, $public_i$ is an indicator variable that takes value 1 if the student comes from a public school and zero otherwise, X_{1i} is a vector of family background variables such as mother's education and family income of student *i*, and X_{2i} is a vector of school characteristics of student *i*. We define public school students as those who took at least two academic years, including the last year, of their high school education in a public school.[13] In Section 4.2.3, we show that our results are robust to the definition of the variable *Public*. In this subsection, we define two variables: (i) *Full Public*, which characterizes students who study their entire high school education in a public institution; and (ii) *Partial Public*, denoting the students who did at least 1 year of their high school education in a public school. We show that both variables have a negative and statistically significant effect on test scores.

Our estimation method is, first, ordinary least squares (*OLS*) where standard errors are robust to the presence of heteroskedasticity. We also analyze quantile regression estimates, first introduced by Koenker and Basset (1978).[14] The least squares estimator specifies and estimates the conditional mean function, $E[Y/X=x]=x\beta$, where Y is a univariate random variable and x is a vector of covariates with the associated parameter vector β. Quantile regression specifies and estimates a family of conditional quantile functions $F^{-1}_{y/x}(\tau/x) = x\beta(\tau)$; where F is the conditional distribution function of Y given X and τ, a quantile in the interval [0,1].[15] Thus, quantile regression provides several summary statistics of the conditional distribution function, rather than just one characteristic, namely, the mean. This descriptive advantage of quantile regression allows us to characterize and distinguish the effects of covariates, for example, the variable *public*, on the upper and lower quantiles of the distribution of the entrance test scores. Furthermore, quantile regression is also a robust technique to outliers in the dependent variable.

Three econometric caveats are important to mention. First, observe that rigourously we must transform the dependent variable, $y=ETS$, to $\ln(y/10-y)$ instead of $\ln(y)$, since the *ETS* is bounded by 0 and 10. Notice, however, that the *OLS* estimator is not equivariant to monotonic transformation and therefore we cannot recover from the transformed model the true effect of each independent variable on the *ETS* (see Papke & Wooldridge, 1996). Nevertheless, due to the equivariance property of the quantiles, estimation of quantile regression function for continuous fractional data is relatively simple (see Machado & Silva, 2006; Powell, 1986), specially when there are no mass point at 0 and 10, which is the case of our sample. Quantitatively, estimations with the transformed variable did not differ significantly from the ones presented here. Since it is easier to interpret the coefficients of the model presented in (1), we just report estimation results for the untransformed model.[16]

In addition, it is difficult to estimate the impact of public school on students' score because there are unobservable characteristics of students and schools that are correlated with each other. Are private schools really better than public ones, or is it simply that better students attend private schools? Here we are not comparing schools among one system, either only public or private schools. It could be the case that the best public (or private) schools, have the best teachers, and also the brightest students. Our goal is to compare the performance of students between a set of public and private schools. It is important to highlight that in Brazil usually poor parents are financially constrained and therefore cannot send their children to a private school. Parents that can send their children to a private school, in general, because of known differences in quality,[17] do so. Therefore, financial issue might be a key constraint in parents' choice of sending their children to a private or a public school in Brazil. However, there might be some exemptions. It might be the case that parents with a very bright child might

[11] Data from the same survey show that doctors make on average four times more than high school teachers.

[12] Lazear (2001) presents a model of the educational production function. See also the recent article by Rivkin, Hanushek, and Kain (2005).

[13] In Brazil the high school education is composed of three academic years.

[14] See Koenker (2005) for more details on quantile regression estimation and inferences. Koenker and Hallock (2001) is an excellent heuristic surveys of quantile regression.

[15] In our case, we have $Q_{\ln(ETS)}(\tau|X)=X'\beta(\tau)$, $\tau \in (0, 1)$, where $X=[1\ public\ X_1 X_2]$. Function $Q_{\ln(ETS)}(\tau|X)$ is the conditional quantile distribution of $\ln(ETS)$ given X and τ.

[16] The results of the transformed model are available upon request.

[17] See, for instance, Menezes-Filho (2007). He shows that not only the proficiency of public school students is lower than private school students, but also the fraction of college educated teachers are lower in public than in private schools. Teachers' wage are also lower in public than private schools.

make an extra effort to send this child to a private school, while parents with children that are struggling at school might decide to send them to a public one even if they could afford to pay a private school.[18] We believe that this effect is marginal, but this could bias the estimated coefficient, β_1, of the variable *public*. Specifically, we might have that the error term, ε, is not orthogonal to the regressors, $X = [1\, public\, X_1 X_2]$. A final difficulty in interpreting estimates of Eq. (1) steams from the fact that not everyone takes this college entrance exams. The data set includes only students that actually sat for the college admission exam. Since the decision to take the exams is likely to be correlated with a student's potential score, the estimates of Eq. (1) would not reflect the true difference among public and private school students. However, since take-up rates are likely to be much larger for private school than public school students, our estimation misses most low performing public school students that never bother to sign-up for the entrance exam. Therefore, in this case if we find that public school students perform worse than private school ones in this entrance exam, in practice our results would underestimate the negative impact of public schools.[19] There are two ways to circumvent the problems raised above. First, if we can control for students ability in regression (1), then the bias might vanish. Alternatively, we can use standard instrumental variable techniques to address the orthogonality problem.

In order to control for students' cognitive ability we use the score at *ENEM*.[20] Contrary to the Entrance Exam (*Vestibular*) which evaluates students in their knowledge of specific subjects (see below the subjects), the *ENEM* exam is more general and emphasizes logical questions. The difficulty in this approach is that this exam is not mandatory[21] and in our sample less than half of the students did this exam.

The potential problem of unobservable selectivity implies that our *OLS* regression might not be capturing the exogenous effects of public schools on test scores. In order to address this problem, the standard approach is to use instrumental variable (IV) techniques. The validity of this procedure, however, depends on the presence of good instruments for the variable *public*, such that: the instrumental variable should be correlated with the variable *public* and should be orthogonal to the error term of the regression equation (1). Since we do not have a valid instrument and it is very hard to address the bias coming from unobservable variables in non-experimental data, we use the technique developed by Altonji et al. (2005). They propose a procedure to investigate the potential size of any bias due to unobservable variables of the estimated coefficient of some specific variable, in our case the variable *public*, due to unobservable selectivity.[22] In their paper they propose the idea that "*selection on observables is the same as selection on unobservables*" (Altonji et al., 2005, p. 154), which, in our case, is equivalent to the condition that $Cov(\varepsilon, Public)/Var(\varepsilon) = Cov(\beta_2 X, Public)/Var(\beta_2 X)$, where X is a vector of observable characteristics, and ε is the error term potentially correlated with *Public*. This is a valid procedure when the point estimates are sensitive to the inclusion of additional control variables, which corresponds to our case, since when we introduce family background variables the estimated coefficient of the variable *Public* decreases in magnitude (see Section 4.1). The biased from *OLS* is $Cov(\varepsilon, \widetilde{Public})/Var(\widetilde{Public})$, where tildes denote the residuals from a regression of *Public* on X, and can be assessed by the following equation[23]:

$$\begin{aligned}\frac{Cov(\varepsilon, \widetilde{Public})}{Var(\widetilde{Public})} &= \frac{Cov(\varepsilon, Public)}{Var(\widetilde{Public})} = \frac{Cov(\varepsilon, Public)}{Cov(\beta_2 X, Public)}\frac{Var(\beta_2 X)}{Var(\varepsilon)} \\ &\quad \times \frac{Cov(\beta_2 X, Public)}{Var(\beta_2 X)}\frac{Var(\varepsilon)}{Var(\widetilde{Public})} \\ &= \frac{Cov(\beta_2 X, Public)}{Var(\beta_2 X)}\frac{Var(\varepsilon)}{Var(\widetilde{Public})},\end{aligned}$$

where the first equality follows if ε and X are orthogonal and the second equality follows from the fact that $Cov(\varepsilon, Public)/Var(\varepsilon) = Cov(\beta_2 X, Public)/Var(\beta_2 X)$. Therefore, instead of trying to find an exogenous variation on the variable *Public*, we will calculate the potential bias coming from unobservable variables to see if results could be different.

3. The data

We use a data set on students entrance test scores at Universidade Federal de Pernambuco (UFPE) which is the major University in Brazilian Northeast. UFPE is a public university in which there is no tuition fees and the main entrance requirement is a test that evaluates students on all subjects of high school years. The students took the entrance test (*vestibular*) in 2005 and they are required to choose their major before they take the exam. The exam evaluates students in many subjects, such as: Mathematics, Biology, Physics, Chemistry, Portuguese, a foreign language (either English, French or Spanish), Literature, Geography, and History. There are two rounds. In the first round, students are evaluated in all subjects and the score in this round is an average of the performance in all subjects. In the second round, the subjects are major specific and the final score is an weighted average of the score in the first and second round. In order to be able to compare all students, we use only the score of the first round.[24] We use two data sets. One comes from COVEST,[25] which is the institution that organizes the entrance exam. In order to sign in for this exam, students must answer a questionnaire with questions about family background, personal and school characteristics.

In order to evaluate the students in the university, we also use data from SIGA,[26] which has information on students' Grade Point Average (GPA). University student record data are rich with detailed information on individual characteristics, family background, the colleges they applied for, and their previous school. Our two main variables are: students' scores on the entrance test, and students' GPA in their first academic year at the university. Scores range from 0 to 10. The explanatory variables can be divided into,

[18] See Neal (2002) for a summary and some references on the effects of public schools on students performance in test scores.

[19] In fact, Emilio et al. (2004) find that the correction for this selection bias has a small impact on the coefficients of a probit model on the determinants of the access to the University of São Paulo.

[20] National Exam for High School Students.

[21] There are some incentives for students to do this exam. There are some scholarship provided by the Brazilian government for public school students to study at private universities (*PROUNI*). Students are selected for this scholarship according to some criteria, and one of them is the score at the *ENEM* exam. Also, some universities (including some public ones) are using the score at *ENEM* to recruit students.

[22] We also tested as an instrument for the dummy variable *public* the indicator variable that takes value 1 if the student's house has more than one air conditioner and takes value zero otherwise. The sign and statistical significance of the variable *public* were robust to such approach. However, this indicator variable might be correlated with unobservable variables that affect the entrance test score such as father's and/or mother's ability. Therefore, we do not report results using such approach, since such instrument might not be valid.

[23] The bias is given by $p\lim \beta_1 = \beta_1 + (Cov(\varepsilon, \widetilde{Public})/Var(\widetilde{Public}))$ and it is positive as long as the variable *Public* is not orthogonal to the error term ε.

[24] Also, students are required to have a minimum grade in the first round to go to the second round.

[25] Commission of Selective and Training Process.

[26] SIGA is the Academic Information System at UFPE.

Table 1
Summary statistics: sample average and sample standard deviation.

Variables	Whole sample		Private		Public	
	μ_X	σ_X	μ_X	σ_X	μ_X	σ_X
Observations	54,877		34,717		20,160	
Entrance test score	4.36	1.38	4.63	1.40	3.88	1.21
First year GPA	7.58	1.38	7.59	1.38	7.57	1.39
Age	20.54	5.40	19.65	4.66	22.11	6.19
Female	0.56	0.49	0.549	0.497	0. 571	0.49
Married	0.07	0.25	0.049	0.21	0.099	0.299
Father education	11.61	4.40	13.04	4.04	9.288	3.93
Mother education	11.78	4.46	13.27	4.08	9.38	3.95
Working father	0.58	0.49	0.67	0.47	0.45	0.49
Working mother	0.50	0.50	0.57	0.49	0.36	0.48
Family income	1620.45	2072.07	2189.49	2374.73	711.64	895.0
Working student	0.27	0.44	0.21	0.41	0.38	0.48
Living with parents	0.80	0.40	0.83	0.37	0.75	0.42
Whites	0.43	0.49	0.48	0.49	0.33	0.47
Asians	0.05	0.21	0.049	0.21	0.043	0.20
Natives	0.02	0.13	0.014	0.12	0.022	0.14
pardos	0.38	0.48	0.33	0.47	0.45	0.49
Blacks	0.09	0.29	0.06	0.23	0.09	0.28
Catholics	0.54	0.50	0.57	0.49	0.48	0.49
Protestants	0.22	0.41	0.16	0.36	0.30	0.46
Jewish	0.002	0.04	0.002	0.046	0.001	0.031
Afro-religion	0.006	0.08	0.006	0.078	0.005	0.073
Other religions	0.09	0.29	0.096	0.29	0.081	0.273
Atheist	0.11	0.31	0.108	0.311	0.112	0.315
Hours worked	1.69	2.92	1.261	2.565	2.43	3.32
Internet user	0.35	0.47	0.479	0.499	0.135	0.34
Reading	0.29	0.45	.222	0.415	0.389	0.487
Lab. classes	0.36	0.48	.462	0.498	0.197	0.397
Foreign language	0.04	0.20	0.059	0.236	0.014	0.118
Private classes	0.40	0.49	0.461	0.498	0.3	0.458
Supletivo	0.03	0.18	0.034	0.181	0.034	0.182
ENEM score	5.79	1.77	6.22	1.70	5.05	1.65

Personal Information such as age, gender, marital status, race, religion, number of children, parents schooling, parents employment status, family income, hours worked and *Academic History* such as school attended (so we can identify the type of school if private or public, catholic or other), if had lab classes, foreign language classes, preparation classes to the entrance test and the college they applied for, among others. It is important to emphasize that our goal is not to measure the impact of public schools on cognitive ability of students. Our goal is to investigate whether there are any barrier to public school students to get in at the university after we control for observed students' characteristics.

3.1. Summary statistics

Table 1 contains summary statistics of the key variables used in this article. Data are from 56,723 students who took the Entrance Test to Universidade Federal de Pernambuco (UFPE) in 2005. Roughly 5400 students are admitted per year. In general, most students that are admitted choose to pursue academic studies at UFPE. We dropped the observations with missing values on test score. This leaves the sample with 54,877 students with roughly equal numbers of males and females. Among those, 34,717 (63%) and 20,160 (37%) come from private and public schools, respectively.

Our sample consists of students 20 years old on average, most of them single and still living with their parents. The majority of our students classify themselves as white or *pardos*[27] (about 81%). Regarding religion, more than 50% of the students are catholics, 21% are protestants, 11% declared themselves as Atheists, and less than 1% are Jewish. Students were also queried about their access to educational resources. In our sample 34% of the students had access to internet, 36% had additional lab classes, and only 4% of the students had foreign languages classes.

The monthly average family income is R$1,620.00 (around 4.5 minimum wages).[28] Notice however that the income distribution across families is very unequal as shown by the standard deviation. In addition, the mean family income is about 3 times higher for private school students than for students that attended public schools. Parents have around 11 years of education, and mothers that send their children to a private school have on average 4 more years of schooling than mothers that send their children to a public school. Almost 60% of their fathers are working while 50% of their mothers have paid jobs. On average 27% of students are working around 1.6 h per day.

In Brazil, there is an alternative way to get a high school degree for students with a large age/grade distortion. This *Supletivo* (Supplementary) method offers short-term courses with a condensed material for different grades. The students can get, for instance, middle school diploma in a 1 year course. In our sample about 3% of students got their secondary school degree with a *Supletivo* method.

The main explanatory variable is the student's achievement on the entrance test, which ranges from 0 to 10. The average test score among all students is 4.36. Fig. 1 shows the empirical density estimate of the entrance test score for students from private and public schools. It shows that public school students score distribution is located to the left of that for private school students. This implies

[27] Due to interbreeding of races (blacks and whites, natives and whites and blacks and natives) individuals classify themselves as brown or *pardos*.

[28] The Real/Dollar exchange rate in the period of when the data was collected was 2.856. This translates into a monthly average family income of US$ 567.22.

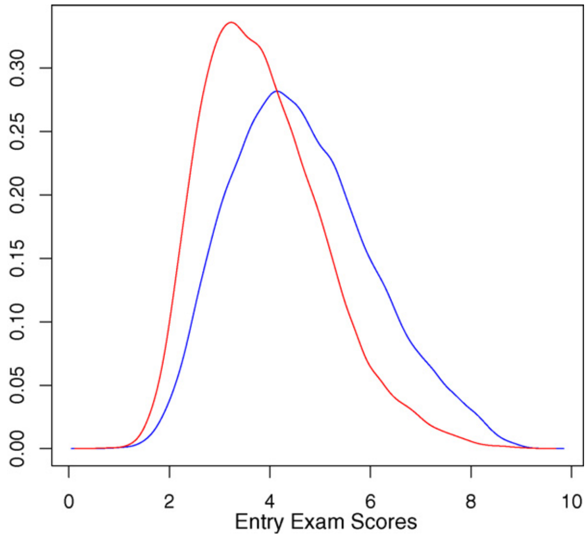

Fig. 1. Empirical density of the entrance test score. Students from private schools: solid blue line. Students from public schools: dotted red line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

that there is a large mass of public school students on low scores than private school students. The average test score is also lower for public school students (3.88) than for private school ones (4.63). In this paper we study the following important questions: Is the difference in the entrance test scores between public and private students due to family background and school characteristics? Or is this difference explained by unobservable effort and ability? Moreover, once the students from public schools get in the university, do they have a performance similar to those that come from private schools?

4. Basic results

4.1. OLS results

Table 2 presents results for the specifications using the *OLS* estimation procedure. The first noticeable result is the negative sign of the coefficient of the indicator variable *public*, which strongly suggests that, public schools have a negative effect on entrance test scores when compared to private schools. In *all* regressions of part I of Table 2, the coefficient of the dummy public is negative and also statistically different from zero at 99% confidence level.[29] The sign and statistical significance of this variable is robust to the introduction of family background variables[30] (column (3)), and school characteristics (column (4)). Observe also that the R-squared indicates that the most complete specification (column (4)) captures about 30% of the total variability in the entrance test scores. In our most parsimonious specification, in column (1), the public school dummy explains about 7% of total variability in score performances. Quantitatively, we have from column (1) that test scores of public school students are on average about 17% lower than those taken by private school students. However, once we control for individual and family characteristics, such as mother's years of schooling, family income, and religion,[31] this effect decreases from 17% to 7% (see column (3)). It decreases further to 6% once we control for another set of variables, such as whether the school has or not lab classes (see column (4)).

In columns (3) and (4), all other additional controls display stability as to the sign, size and significance of the associated regression coefficient, except the variable age, number of children, family income, and the dummy for Asian students. Regarding family background, we observe that mother's years of schooling is an important determinant of the entrance test score. This is consistent with two recent studies (Bourguignon et al., 2003; Ferreira & Veloso, 2006), which show, using household survey data, that parent education is a strong predictor of schooling and intergenerational educational persistence in Brazil is high by international standards. Quantitatively, we also observe the strong impact of the variable foreign language on test scores. The fact that students studied a foreign language increases their scores in about 10%. Foreign language might indicate student's ability and learning capabilities, but it is important to highlight that in the entrance exam students are also evaluated on either English or Spanish.

Interestingly, the dummy for races are negative for native Brazilians and blacks. On average native Brazilians and blacks score about 7.7% and 2.7% lower than whites, respectively, after we control for family background, individual and school characteristics. This is particularly important since some universities in Brazil have recently implemented some affirmative policies and introduced quotas for black and native Brazilian students. Observe that the dummy for Asian descendent is also negative and statistically significant once we control for school characteristics (columns (4) and (5)). In addition, notice that *Pardos* score about 0.6% higher than whites and this coefficient is statistically significant in all regressions.

In Table 2 part II, we address any attrition problem. In our most complete specification (column (4)) we miss about 22% of observations relatively to the more parsimonious model. Therefore, in order to evaluate how the coefficient change when we include more controls, we estimate specifications (2) and (3) using the same observations as in column (4). Estimated coefficients are reported in Table 2, columns (5) and (6). We notice that the coefficient of the indicator variable *public* does not change in sign, magnitude (the absolute value of this coefficient increases in column (6) when compared to column (3)), and statistical significance.

Table 2 shows that the magnitude of the point estimate of the variable *public* are sensitive to the inclusion of observable regressors, ranging from −18.9% in models with no control to −6% in models that include detailed controls. Private schools are therefore apparently more advantaged on the basis of observable characteristics that positively affect academic performance. It is not clear, however, if private school students have also advantages in terms of unobservable characteristics (i.e., motivation, effort, cognitive ability) that are correlated with test scores and are not controlled for in the regressions presented so far. It is important to recall also that we are considering only students that are taking exams and we might be missing low performing students that do not sign up to take this exam, and the share of such students might be larger in public than in private schools.

The last row of Table 2 reports the estimated bias in the most complete specification (column (4)), using Altonji et al. (2005) procedure. The bias is equal to −0.018, which is an evidence of a negative correlation of unobservable variables in the test

[29] The *t*-critical value for 99% confidence level is 2.58.

[30] The difference of column (1) and (2) is that in column (2) we control for whether the public school is a federal or a military School. There are two federal schools in the sample and one military school. Federal and military schools have in general good scores in standardized test. Although they are public there is a very competitive entrance exam in those schools, which bias the sample. In fact, observe that once we control for those schools the absolute value of the public dummy increases.

[31] We include five indicator variables for religion and the Catholic religion is the reference group. They are: Protestant, Jewish, Atheist, Afro-Religions, and others.

Table 2
Determinants of the entrance test score – *OLS* estimation.

	Part I				Part II, attrition	
	(1)	(2)	(3)	(4)	(5)	(6)
Public	−0.176**	−0.189**	−0.072**	−0.060**	−0.189**	−0.089**
	(−10.63)	(−12.93)	(−10.31)	(−8.99)	(−12.60)	(−10.20)
Age			0.002*	−0.011**		0.001*
			(1.93)	(−11.56)		(1.71)
Married			0.010	0.030**		0.008
			(0.79)	(4.55)		(0.99)
Gender (female = 1)			−0.036**	−0.049**		−0.034
			(−7.23)	(−9.73)		(−7.31)
Number of children			−0.044**	0.016**		−0.044
			(−7.36)	(3.42)		(−7.55)
Hours worked			−0.007**	−0.009**		−0.007
			(−8.16)	(−13.05)		(−8.15)
Asian			−0.003	−0.013*		−0.005
			(−0.59)	(−1.94)		(−0.94)
Natives			−0.077**	−0.077**		−0.079**
			(−7.94)	(−7.84)		(−8.08)
Pardos			0.008**	0.008**		0.008**
			(2.93)	(2.60)		(2.53)
Blacks			−0.022**	−0.019**		−0.022**
			(−3.45)	(−3.59)		(−3.92)
Living with parents			−0.041**	−0.033**		−0.042**
			(−7.49)	(−9.79)		(−7.91)
Religion	No	No	Yes	Yes	No	Yes
Mother's schooling			0.008**	0.004**		0.007**
			(13.27)	(7.91)		(13.99)
Father's schooling			0.007**	0.004**		0.006**
			(7.92)	(6.38)		(7.29)
Working mother			0.005	0.006**		0.006
			(1.55)	(2.00)		(1.47)
Working father			0.016**	0.009**		0.013**
			(6.29)	(4.23)		(5.61)
Monthly family income			2.3e$^{-0.5}$**	1.7e^{-05}**		2.3e^{-05}**
			(6.15)	(6.00)		(6.27)
Internet user				0.058**		
				(10.48)		
Lab classes				0.035**		
				(4.35)		
Foreign language				0.104**		
				(11.92)		
Reading habit (yes = 1)				0.059**		
				(20.77)		
Supletivo				−0.101**		
				(−5.00)		
Number of tests taken				0.101**		
				(18.59)		
Private preparation classes				0.065**		
				(10.44)		
Non-12th grade students				−0.030*		
				(−1.79)		
Federal and military schools	No	Yes	Yes	Yes	Yes	Yes
N. of observ.	54877	54877	44782	42615	42615	42615
Adjusted R^2	0.068	0.088	0.158	0.294	0.089	0.174
Estimated bias				−0.018		

All specifications include a constant, not reported. *T*-statistics are presented in parentheses, using heteroskedasticity-consistent standard errors clustered at the school level.
* Significant at the 90% confidence level.
** Significant at the 95% confidence level.

score equation and the variable *Public*, which suggests that the estimated effect of public schools on student's performance is overestimated.[32] However, our estimated coefficient is −0.06, then with a bias of −0.018, the lower bound coefficient of the variable *Public* would be −0.042. In addition, the ratio of the coefficient of the variable *Public* and the estimated bias is:

$$\text{ratio} = \frac{-0.06}{-0.018} = 3.333.$$

This implies that the effects of unobservable variables on the estimated coefficient of the variable *Public* would have to be 3.333 times stronger than the effect of the observable variables on this same coefficient in order to explain the entire negative effect of the variable *Public* on test scores. We believe that this is rather unlikely.

4.2. Robustness

In this subsection, we perform several robustness exercises. In the first part, we control for students' cognitive ability and use the score at the *ENEM* (National Exam for High School students) as an independent variable in the model. As we argued previously, the

32 Given the assumption that the selection of observables is similar to the selection of unobservables, we just need to check this for the most complete specification.

Table 3
Summary statistics: sample average and sample standard deviation.

Variables	Whole sample		*ENEM*		No *ENEM*	
	μ_X	σ_X	μ_X	σ_X	μ_X	σ_X
Observations	54,877		23,807		31,070	
Public	0.37	0.48	0.36	0.48	0.37	0.48
Entrance test score	4.36	1.38	4.74	1.47	4.06	1.23
Age	20.54	5.40	18.92	2.99	21.78	6.41
Female	0.56	0.49	0.60	0.49	0.52	0.50
Married	0.07	0.25	0.02	0.16	0.10	0.30
Father education	11.61	4.40	11.87	4.41	11.40	4.38
Mother education	11.78	4.46	12.15	4.46	11.50	4.43
Working father	0.58	0.49	0.64	0.48	0.55	0.50
Working mother	0.50	0.50	0.53	0.50	0.47	0.50
Family income	1620.45	2072.07	1705.42	2185.36	1556.56	1980.248
Working student	0.27	0.44	0.21	0.44	0.28	0.41
Living with parents	0.80	0.40	0.85	0.40	0.76	0.42
Whites	0.43	0.49	0.44	0.49	0.43	0.49
Asians	0.05	0.21	0.05	0.21	0.05	0.21
Natives	0.02	0.13	0.02	0.12	0.02	0.14
pardos	0.38	0.48	0.39	0.48	0.38	0.48
Blacks	0.09	0.29	0.09	0.28	0.09	0.29
Catholics	0.54	0.50	0.59	0.49	0.56	0.49
Protestants	0.22	0.41	0.21	0.41	0.22	0.41
Jewish	0.002	0.04	0.001	0.04	0.002	0.04
Afro-religion	0.006	0.08	0.004	0.07	0.007	0.08
Other religions	0.09	0.29	0.08	0.27	0.09	0.29
Atheist	0.11	0.31	0.11	0.31	0.11	0.31
Hours worked	1.69	2.92	0.91	2.21	2.26	3.23
Internet user	0.35	0.47	0.37	0.48	0.33	0.47
Reading	0.29	0.45	0.29	0.45	.28	0.45
Lab. classes	0.36	0.48	0.37	0.48	.462	0.498
Foreign language	0.04	0.20	0.04	0.20	0.04	0.20
Private classes	0.40	0.49	0.45	0.49	0.36	0.48
Supletivo	0.03	0.18	0.01	0.10	0.05	0.21

ENEM exam is more general than the *Vestibular* and it also evaluates students in logical questions. A caveat must be added, however. This exam is not mandatory and this might bias our results. In the second part, we control for city fixed effects and school fixed effects. The third part investigate how other definitions of our variable *Public* affect test scores.

4.2.1. *Cognitive ability (ENEM exam)*

It is important to emphasize that the *ENEM* score already absorbs good variation in characteristics (observable and non-observable) that affect students academic performance. Therefore, the control for the *ENEM* score is a demanding exercise to investigate whether or not public schools impact negatively on students performance at public university entrance exams. The problem of the *ENEM* exam, as we mentioned previously, is that less than half of the students that did the entrance test exam did the *ENEM*. This can bias the results once we control for the *ENEM*, since the students that did this test were not randomly assigned. Table 3 provides summary statistics in which we can compare statistics of the students who did take and who did not take the *ENEM*.[33] Indeed, there are important differences such as the level of monthly income and the entrance test score,[34] which are larger for those that took the *ENEM* exam than those that did not. Interestingly, however, the fraction of public school students in the two samples are roughly the same and the Mann–Whitney test does not reject the null hypothesis of equality of distributions. Therefore, it seems that better students with a more favorable family background took the *ENEM* exam, but the distribution of public school students of those that did take and did not take the exam are not statistically different.[35] Therefore, it is not straightforward to identify the direction of the bias of the variable *public* due to the no random choice of taking or not the *ENEM*.

With this caveat in mind, Table 4, Part I, contains the estimated parameters of our model when we introduce the score at the *ENEM* exam as an independent variable to control for any omission in cognitive ability. We estimate similar regression to those in columns (2)–(4) of Table 2, but we add the *ENEM* student's score as an explanatory variable. Parameters are reported in columns (1)–(3) of Table 4. Since we lose about half of the observations when we add the *ENEM* score, for comparison reasons we also estimate in part II of Table 4 the model without the *ENEM* score (columns (4)–(6)), but with the same observations of columns (1)–(3). Notice that the model with the variable *ENEM* explains about 70% of the total variability in the entrance test scores, while the model with the most complete specification, but without the *ENEM* control explains about 36% of such variability.

Observe that the magnitude of the coefficient of the variable *public* decreases in magnitude, but it is still negative and statistically different from zero (at 99% confidence level) once we add the *ENEM* student's score as a regressor.[36] Test scores of public school students are on average about 5.6–3.7% lower than those taken by private school students once we control for the *ENEM* scores, and depending whether or not we control for family background and school characteristics. Notice that contrary to other variables,

[33] We also run a probit model to study the determinants of taking the *ENEM*. For the sake of space, we do not report results here.
[34] As well as mother's and father's education, among other variables.
[35] In fact, the entrance test score for those that took the *ENEM* is large than those that did not do the *ENEM* for both public and private school students.
[36] Compare the first row of columns (1)–(3) with the first row of columns (4)–(6).

Table 4
Determinants of the entrance test score – *ENEM* exam.

	Part I			Part II		
	(1)	(2)	(3)	(4)	(5)	(6)
Public	−0.056**	−0.042**	−0.037**	−0.246**	−0.135**	−0.085**
	(−12.38)	(−9.07)	(−7.09)	(−10.96)	(−11.31)	(−8.35)
Age		0.001*	−0.006**		0.011**	−0.018**
		(1.89)	(−9.55)		(7.47)	(−11.96)
Married		−0.006	0.006		0.005	0.045**
		(−0.64)	(0.62)		(0.38)	(2.99)
Gender (female = 1)		0.026**	0.017**		−0.046**	−0.059**
		(7.75)	(4.20)		(−7.37)	(−9.05)
Number of children		−0.032**	0.005		−0.089**	0.053**
		(−5.76)	(1.01)		(−7.75)	(5.68)
Hours worked		−0.007**	−0.005**		−0.010**	−0.009
		(−9.89)	(−10.27)		(−9.10)	(−10.89)
Asian		0.006	0.003		−0.004	−0.018*
		(1.62)	(0.66)		(−0.53)	(−1.65)
Natives		−0.002	−0.004		−0.069**	−0.065**
		(−0.26)	(−0.50)		(−3.47)	(−4.27)
Pardos		0.002	0.001		0.001	−0.001**
		(0.96)	(0.09)		(0.01)	(−0.23)
Blacks		−0.001	−0.002		−0.025**	−0.020**
		(−0.35)	(−0.63)		(−3.32)	(−2.74)
Living with parents		−0.014**	−0.011**		−0.040**	−0.027**
		(−4.74)	(−3.76)		(−6.29)	(−5.41)
Religion	No	Yes	Yes	No	Yes	Yes
Mother's schooling		0.001*	0.001		0.007**	0.004**
		(3.28)	(1.21)		(13.19)	(7.42)
Father's schooling		0.001**	0.001**		0.008**	0.005**
		(2.68)	(3.32)		(11.89)	(10.19)
Working mother		0.001	0.002		0.009	0.007**
		(0.66)	(0.84)		(1.64)	(1.99)
Working father		0.005**	0.003		0.012**	0.005
		(2.68)	(1.81)		(3.55)	(1.52)
Monthly family income		3.1e$^{-0.6}$**	3.3e$^{-0.6}$**		2.3e$^{-0.5}$**	1.6e$^{-0.5}$**
		(3.22)	(3.64)		(6.74)	(6.39)
Internet user			0.002			0.051**
			(0.74)			(8.30)
Lab classes			0.001			0.031**
			(0.37)			(3.63)
Foreign language			0.013*			0.080**
			(1.96)			(7.54)
Reading habit (yes = 1)			0.025			0.053**
			(11.94)			(14.00)
Supletivo			−0.034*			−0.072**
			(−3.58)			(−5.70)
Number of tests taken			0.030**			0.120**
			(8.85)			(19.04)
Private preparation classes			0.035**			0.074**
			(10.41)			(14.87)
Non-12th grade students			−0.059**			−0.040
			(−3.22)			(−0.94)
ENEM score	0.149**	0.147**	0.139**			
	(94.78)	(92.18)	(152.09)			
Federal and military schools	Yes	Yes	Yes	Yes	Yes	Yes
N. of observ.	23807	20277	19464	23807	20277	19464
Adjusted R^2	0.714	0.716	0.726	0.144	0.239	0.364
Estimated bias			−0.012			

All specifications include a constant, not reported. *T*-statistics are presented in parentheses, using heteroskedasticity-consistent standard errors clustered at the school level. (*) and (**) indicate same confidence level as in Table 2.

such as race, gender, mother's schooling, among others, the sign and statistical significance of the variable *public* are robust to the introduction of *ENEM* test score.[37]

4.2.2. City fixed effects and school fixed effects

Here we use the most complete specification reported on column (4) of Table 2 to investigate the effects of city and school fixed effects. The first part of Table 5 reports the estimated coefficient of the variable *Public* when we control for city and school fixed effects. For the sake of space we do not report the estimated coefficients of the other variables. The introduction of city fixed effects might be important for the variable *Public* because in small cities there are less choice for students in terms of schools than in large cities. In addition, it might be the case that there are more public schools relative to private schools in small cities than in large cities and if schools in small cities have less quality than in large cities, the variable *Public* might be capturing a "small city" effect. We observe that the estimated coefficient, which captures the effects of public schools on scores, is most unchanged when we introduce city fixed effects (Table 5, row (1)). It decreases in absolute value from

[37] For instance, the coefficient of the variables blacks and natives are not statistically different from zero when we add the variable *ENEM*.

Table 5
Determinants of the entrance test score – city and school fixed effects.

	Coeff. on *Public*	Robust *t*-value	City fixed effects	School fixed effects	Estimated bias	Ratio
(1)	−0.057**	−16.52	Yes	No	−0.015	3.8
(2)	−0.023**	−3.52	No	Yes	−0.010	2.3
(3)	−0.021**	−3.13	Yes	Yes	−0.009	2.333

All specifications include a constant and the variables used in column (4) of Table 2. (*) and (**) indicate same confidence level as in Table 2.

Table 6
Determinants of the entrance test score – definition of the variable *Public*.

	Coeff. on full *Public*	Robust *t*-value	Coeff. on partial *Public*	Robust *t*-value	Number of observations	Adjusted R^2
(1)	−0.071**	−10.97	−0.047**	−8.82	42,587	0.29

All specifications include a constant and the variables used in column (4) of Table 2. (*) and (**) indicate same confidence level as in Table 2.

6% to 5.7% and it is also statistically different from zero at usual confidence level.

We also introduce school fixed effects to get rid of sorting problems associated with different qualities of schools, different neighborhoods and parental backgrounds This is also a demanding exercise, since this would capture schools effects, such as the way the schools are run and the quality of teachers, which might explain difference in scores between private and public school students; this would also capture neighborhood effects, such as crime rate in the school's neighborhood and average academic performance among schools. Interestingly, we observe that the estimated coefficient for the variable *Public* decreases in magnitude from 6% to 2.1% when we introduce school and city fixed effects, but it is still negative and statistically different from zero at usual confidence intervals (Table 5, rows (2) and (3)).

We also estimated the bias of the variable *Public* using the method proposed by Altonji et al. (2005). We found that the effect of unobservable variables would have to be at least 2.3 larger than the effect of all observed variables, including city and school fixed effects, in order for the effect of the variable *Public* on student's score to vanish. We believe that the results above corroborated our previous finding that public schools have a negative effect on student's score, even when we control for family background and school characteristics.

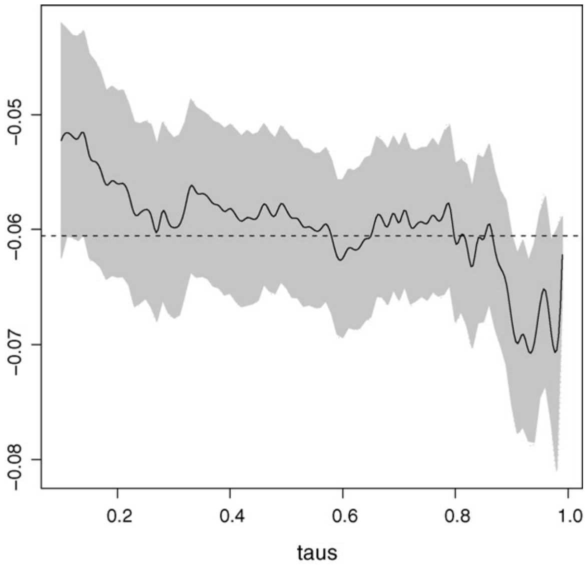

Fig. 2. Quantile regression estimates for the indicator variable public. Dependent variable: entrance test score. The solid line corresponds to the quantile estimate and the shaded area the 95% confidence interval. The dotted line represents the *OLS* estimate.

4.2.3. Definition of the variable Public

Finally, we change the criterion of the definition of the variable *Public*. Previously, we defined public school students as those who took at least 2 years, including the last year, of their high school education in a public school. Now, we define two variables: (i) *Full Public*, which characterizes students who study their entire high school education in a public institution; and (ii) *Partial Public*, denoting the students who did at least 1 year of their high school education in a public school. Results are reported on the last row of Table 5. We observe that both coefficients are negatively and statistically different from zero at 99% confidence level. The coefficient of the variable *Full Public* is about 1.51 larger than the coefficient of the variable *Partial Public*. Therefore, results seem to be robust to the definition of our variable public school (Table 6).[38]

4.3. Quantile results

In this subsection we analyze the quantile regression estimates. Since quantile regression procedures produce one point estimation for each quantile, for the sake of space, we focus only on the coefficient of the indicator variable *public*.[39] We use the same variables used in column (4) of Table 2, and results are displayed in Fig. 2. This plot shows the quantile regression (QR) estimates as well as the 95% confidence intervals. For comparison reasons, the least square estimate is presented by the dotted horizontal line.

Quantile regression provides the appropriate tool to determine whether there are any difference in marginal responses of the entrance test score to different school systems (public or private). Observe that the coefficient of the dummy variable *public* is negative for all quantiles (including the confidence interval) and it varies from −5.2% to −7% as we move from low to high quantiles of the conditional Test Score distribution. Therefore, not only students from public schools have on average lower scores on the Entrance Test Exam, but also students that have higher conditional scores suffer from a greater penalty for attending public schools compared to low conditional score students. We may conclude that students with high performance in the entrance test scores would benefit most by moving from public to private schools.

[38] For this regression, we do not calculate the bias of the estimated public school effects on test scores because the procedure developed by Altonji et al. (2005) considers only one endogenous variable.

[39] Upon request we provide the quantile estimation for the other parameters.

Table 7
Determinants of first semester GPA – *OLS* estimation.

	(1)	(2)	(3)	(4)	(5)
Public	−0.005	0.018*	0.018*	0.009	0.007
	(−0.59)	(1.82)	(1.73)	(1.01)	(0.82)
ETS		0.042**	0.051**	0.063**	0.0611**
		(9.63)	(9.76)	(12.48)	(11.89)
Age			−0.0002	−0.00051	−0.0002
			(−0.55)	(−1.36)	(−0.69)
Married			0.0373	0.037	0.059
			(0.96)	(1.06)	(2.52)
Gender (female = 1)			0.124**	0.057**	0.058**
			(13.51)	(7.85)	(7.69)
Number of children			0.028*	0.020	0.013
			(1.66)	(1.28)	(0.91)
Hours worked			−0.002*	−0.004*	−0.003*
			(−1.22)	(−2.41)	(−1.84)
Asian			−0.023	−0.003	0.0008
			(−1.25)	(−0.23)	(0.05)
Natives			−0.007	−0.004	−0.002
			(−0.19)	(0.14)	(−0.08)
Pardos			0.0068	0.0065	0.008
			(0.73)	(0.78)	(0.96)
Blacks			−0.007	−0.004	−0.001
			(−0.42)	(−0.28)	(−0.10)
Living with parents			−0.0035	0.004	0.00067
			(−0.28)	(0.44)	(0.06)
Religion			Yes	Yes	Yes
Mother's schooling			−0.0007	0.0002	0.0004
			(−0.58)	(0.23)	(0.41)
Working mother			0.005	−0.0033	−0.0011
			(0.55)	(−0.42)	(−0.15)
Monthly family income			−1.08e^{-06}	−7.93e^{-07}	−1.44e^{-06}
			(−0.59)	(−0.48)	(−0.88)
Internet user					0.003
					(0.40)
Lab classes					0.008
					(1.10)
Foreign language					0.0006
					(0.05)
Reading habit (yes = 1)					0.012
					(1.58)
Supletivo					−0.018
					(−0.71)
Number of tests taken					−0.018**
					(−4.17)
Colleges (Medicine, sciences, and humanity)	No	No	No	Yes	Yes
N. of observ.	3661	3660	3166	3166	3087
R^2	0.0001	0.0271	0.0926	0.2973	0.3012

All specifications include a constant, not reported. *T*-statistics are presented in parentheses, using heteroskedasticity-consistent standard errors.
* Significant at the 90% confidence level.
** Significant at the 95% confidence level.

4.4. How is the performance of public school students in the university?

In this subsection we investigate the performance of public school students once they get into the university. Do public school students perform worse than private school students in the university? It might be that only high unobservable ability (e.g., cognitive learning ability and effort in academic studies) students from public schools are approved in the entrance exam and therefore they might perform better than private school students. Observe that on average public school students have a GPA in the first academic year that is similar to that of private school students (see Table 1). The standard deviation is also similar between the two groups. However, it is important to emphasize that different colleges have different ways to evaluate students, and difference in GPAs between public and private school students might reflect differences in the percent of public school students in each college. We therefore control for the colleges that students are enrolled, since, for instance, medical students take a completely different set of courses than economics students. Table 7 provides *OLS* estimation for an equation similar to (1), but in which the dependent variable is the first year ln(*GPA*) of each student.

We observe that the coefficient of the dummy *public* is negative but it is not statistically different from zero when this variable is the only regressor (column (1)). Once we control for the entrance test score in the university (column (2)), the coefficient of the dummy *public* becomes positive and statistically different from zero at 90% confidence level. Students that come from public schools in this case have a GPA in the first year that is about 1.8% higher than students from private schools. This coefficient is robust to the introduction of individual and family characteristics (column (3)), but becomes statistically insignificant once we control for each college (column (4)), and school characteristics (column (5)). In our most complete specification, the regressors explain roughly 30% of the observed variability in the first academic year GPA.

Regarding the other controls, we observe that the entrance test score is the most robust predictor of academic performance in the university. This is expected, since the entrance test score reflects

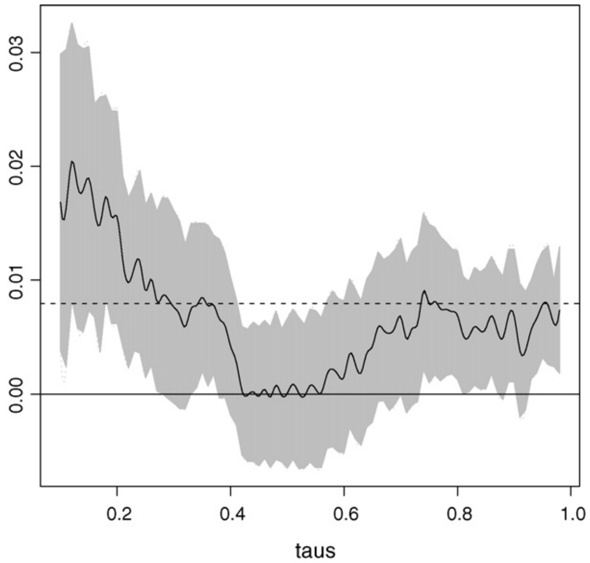

Fig. 3. Quantile regression estimates for the indicator variable public. Dependent variable: first year GPA. The solid line corresponds to the quantile estimate and the shaded area the 95% confidence interval. The dotted line represents the *OLS* estimate.

not only cognitive ability and effort, but also the quality of the education acquired in previous years. Notice that, given the *ETS*, race and family background (as well as school characteristics) have no effect whatsoever on GPA. Interestingly, women have a lower entrance test score than men and this effect is statistically significant in all regressions (see Table 2), but they have a better performance than men in the university and this is robust in all specifications (see Table 7). This is certainly an important question, but we leave it for future research.

We also evaluate how the variable *public* affects the GPA in all quantiles of the conditional distribution. Fig. 3 displays the quantile regression estimates of the coefficient of the variable *public* for the specification defined in column (5) of Table 7. We again focus only on the variable *public*, but the estimated coefficient for the other variables are readily available upon request. Observe that the point estimation of the coefficient of the variable *public* is positive in all quantiles, except for the quantiles close to the median, in which it is roughly zero. For low and high quantiles of the conditional GPA distribution, students from public schools perform better than students from private schools after we control for individual characteristics, family background, entrance test scores, and colleges. This effect is statistically significant at 95% confidence level.[40] At low quantiles of the conditional GPA distribution, the quantile estimated coefficient of the variable public is higher than the one estimated by least squares procedures. However, at high quantiles the estimated coefficient is lower when estimated by quantile regression than the *OLS* estimated coefficient.

The results above suggest that once students get into the university, those from public schools perform better than (or as well as) those from private schools. In addition, the proportion of public school students that gets into the university is about 40%, which is higher than the proportion of students doing the entrance exam at Universidade Federal de Pernambuco (about 37%, see Table 1). However, the results above hide key differences between the two group of students in the university. There is a strong barrier for public school students to get into the university, mainly in high competitive majors, since they have a lower entrance score, even when we control for family background, school characteristics, omitted variables, and unobservable selectivity. The fraction of students from public schools that gets into high competitive majors such as journalism, law, medicine, and electronic engineering is almost null[41] (see Table 8). For instance, in medicine only 6% of the students are from public schools. This number is equal to 7 and 5 for law and electronic engineering, respectively. In journalism there is not only one student from a public school. On the other hand, for less competitive majors the percent of students that came from public schools is much higher. It is, for instance, 100% in music, 65% in geography (teacher credential), 56% in history (teacher credential), and 59% in chemistry (teacher credential). This is an important result, since difference in occupation is an important source of income inequality. Just to give some examples, data from the 2005 Brazilian National Household Sample Survey (PNAD), show that electrical and doctors make about 3 and 4 times more, respectively, than domestic economists and high school teachers.[42] This corroborates the view that the public university educational system in Brazil is a strong barrier to social mobility.

Table 8
Percent of students from public schools by major.

Major	%	Major	%
Journalism	0	Music (chant)	100
Biomedical Engineering	0	Biological Engineering (teacher credential)	68
Publicity/Propaganda	3	Music (instrument)	67
Physics (B.A.)	3	Geography (teacher credential)	64
Production Engineering	3	Chemistry (teacher credential)	59
Electrical Engineering	5	Secretary Studies	55
Computer Engineering	6	History (teacher credential)	55
Medicine	6	Social Science	51
Law	7	Domestic Economics	50

5. Concluding remarks

In this paper we study whether or not the Brazilian high educational system protects the *status quo* by preventing poor families to step up in the social economic ladder. In our knowledge, for the first time we quantitatively estimate the barriers of public school students to get in public universities in Brazil. We use a novel data set on students entrance test scores at Universidade Federal de Pernambuco (UFPE) to evaluate quantitative differences in scores between public and private school students. This is an important factor in the study of inequality, since researchers have shown that test scores that evaluate cognitive skills are key predictor of future wages, and this effect has being growing in recent years due to skill biased technological change. Also, this is a major university in Brazilian Northeast and students do not pay any tuition fee to study at this public university. The dataset has information on individual and school characteristics, and family background.

We found that test scores of public school students are on average about 4.2–17% lower than those taken by private school students, after we control for individual, family, and/or school characteristics, as well as unobservable selectivity. Therefore, public school students not only have lower scores on the entrance because of their "*weaker background*" when compared to private school stu-

[40] Observe that the lower bound of the confidence interval is above zero from 0 to 0.4 quantiles, and from 0.75 to 1 quantiles.

[41] It is important to clarify that law and medicine are undergraduate courses in Brazil. In addition, students choose their major before they take the entrance exam.

[42] Some law professionals (e.g., public attorneys) make on average 10 times more than high school teachers.

dents but also because of the low quality of public schools[43] or lack of preparation for the entrance test exam. Using quantile regression techniques, we also show that students with high performance in the entrance test scores would benefit most by moving from public to private schools. Another important result is that family background (e.g., mother's years of schooling and household's income) is a key predictor of test scores. Finally, we show that once students get into the university, those from public schools perform as well as those from private schools. In addition, the proportion of public school students that gets into the university is roughly the same as the proportion of students doing the entrance exam. However, there is a strong barrier for public school students to get into high competitive majors, such as law, medicine, and electronic engineering is almost null. Our results are therefore consistent with the political economy view (e.g., Fernandez & Rogerson, 1995), which suggests that public university education involves a transfer of resources toward higher income families which in turn strength the persistence of inequality and decrease social mobility. Therefore, policies toward to improve the quality of education in public schools by providing a better preparation for the entrance exam or policies that facilitate access for public school students to get into public universities would improve the equality of opportunities in education for students to acquire a university degree.

References

Alesina, A., DiTella, R., & MacCulloch, R. (2004). Inequality and happiness: Are Europeans and Americans different? *Journal of Public Economics*, *88*, 2009–2042.

Alexopoulos, J., & Cavalcanti, T. (in press). Cheap Home Goods and Persistent Inequality. *Economic Theory*.

Altonji, J. G., Elder, T., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, *113*(1), 151–184.

Altonji, J. G., Elder, T., & Taber, C. R. (2008). Using selection on observed variables to assess bias from unobservables when evaluating Swan-Ganz catheterization. *American Economic Review: Papers & Proceedings*, *98*(2), 345–350.

Barros, R., Carvalho, M., Franco, S., & Mendonça, R. (2009). Determinantes da Queda na Desigualdade de Renda no Brasil. In *Texto para Discussão 1460, IPEA*.

Bellows, J., & Miguel, E. (2009). War and local collective action in Sierra Leone. *Journal of Public Economics*, *93*(11–12), 1144–1157.

Bourguignon, F., Ferreira, F., & Menéndez, M. (2003). Inequality of outcomes and inequality of opportunities in Brazil. *World Bank Policy Research Working Paper 3174*.

Curi, A. Z., & de Menezes-Filho, N. A. (2005). *A Relação entre Desempenho Escolar e os Salários no Brasil*. Mimeo, Universidade de São Paulo.

Doepke, M. (2004). Accounting for fertility decline during the transition to growth. *Journal of Economic Growth*, *9*(3), 347–383.

Emilio, D. R., Belluzo, W., Jr., & Alves, D. C. O. (2004). Uma Análise Econométrica dos Determinantes do Acesso à Universidade de São Paulo. *Pesquisa e Planejamento Econômico*, *34*(2), 275–305.

Fernandez, R., & Rogerson, R. (1995). On the political economy of education subsidies. *Journal of Political Economy*, *62*, 249–262.

Ferreira, S. G., & Veloso, F. A. (2006). Intergenerational mobility of wages in Brazil. *Brazilian Review of Econometrics*, *26*(2), 181–212.

Fuchs, T., & Woessmann, L. (2007). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics*, *32*(2–3), 433–464.

Galor, O., & Zeira, J. (1993). Income distribution and macroeconomics. *Review of Economics Studies*, *60*, 35–52.

Goyal, S. (2009). Inside the house of learning: The relative performance of public and private schools in Orissa. *Education Economics*, *17*(3), 315–327.

Hanushek, E. A. (1995). Interpreting recent research on schooling in developing countries. *The World Bank Research Observer*, *10*(2), 227–246.

Kingdon, G., & Teal, F. (2010). Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach. *Journal of Development Economics*, *91*(2), 278–288.

Koenker, R. (2005). *Quantile regression, Econometric Society monographs*. Cambridge University Press.

Koenker, R., & Basset, G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Koenker, R., & Hallock, K. (2001). Quantile regression: An introduction. *Journal of Economic Perspectives*, *15*, 143–156.

Lazear, E. (2001). Educational production. *Quarterly Journal of Economics*, *CXVI*(3), 777–803.

Machado, J. A. F., & Silva, J. M. S. (2006). *Quantiles for fractional data*. Mimeo, Universidade Nova de Lisboa.

McEwan, P. J. (2008). Can schools reduce the indigenous test score gap? Evidence from Chile. *Journal of Development Studies*, *44*(10), 1506–1530.

Menezes-Filho, N. (2007). *Os Determinantes do Desempenho Escolar do Brasil*. Mimeo, Instituto Futuro Brasil.

Menezes-Filho, N., Fernandes, R., & Picchetti, P. (2006). Rising human capital but constant inequality: The education composition effect in Brazil. *Revista Brasileira de Economia*, *60*(4), 407–424.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skill in wage determination. *Review of Economics and Statistics*, *77*(2), 251–266.

Murphy, K. M., & Peltzman, S. (1997). School performance and the youth labor market. *Journal of Labor Economics*, *22*(2), 299–325.

Neal, D. (2002). How vouchers could change the market for education. *Journal of Economic Perspectives*, *16*(4), 25–44.

Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fraction response variable with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, *11*, 619–632.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, *32*, 143–155.

Psacharopoulos, G. (1999). Returns to investment in education: A global update. *World Development*, *22*(9), 1325-l343.

Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics*, *12*(2), 111–134.

Resende, M., & Wyllie, R. (2006). Retornos para Educação no Brasil: Evidências Empíricas Adicionais. *Economia Aplicada*, *10*(3), 349–365.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Waltenberg, F. D, & Vandenberghe, V. (2007). What does it take to achive equality of opportunity in education? An empirical investigation based on Brazilian data. *Economics of Education Review*, *26*, 710–724.

World Bank. (2007). *World development indicators*. Washington, DC: World Bank.

[43] See Hanushek (1995) for some explanation of differences in quality of public and private schools in developing country, as well as some policy that can reduce such gap in quality.