Department of Economics- FEA/USP

School Reputation and School Choice in Brazil: a Regression Discontinuity Design

ANDREA LÉPINE

WORKING PAPER SERIES Nº 2015-38

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2015-38

School Reputation and School Choice in Brazil: a Regression Discontinuity Design

Andrea Lépine (alepine@usp.br)

Abstract:

The provision of information on schools' performance on standardized tests is expected to generate pressure on schools as students and their families can compare them and make more informed school choices. This paper uses administrative data from Brazil to explore whether the publication of grades obtained at a standardized high school test (the Enem) resulted in changes in enrollments in high and low performing schools, through a sharp regression discontinuity design. The results show that the disclosure of school grades did not result in students reallocating between both types of school, in neither private nor public schools. The findings remain unchanged when I control for the degree of competition faced by schools or their socio-economic environment.

Keywords: School choice, Standardized tests, School accountability

JEL Codes: I25, I28, D83

School Reputation and School Choice in Brazil: a Regression Discontinuity Design *

Andrea Lépine
Universidade de São Paulo

October 2015

Abstract

The provision of information on schools' performance on standardized tests is expected to generate pressure on schools as students and their families can compare them and make more informed school choices. This paper uses administrative data from Brazil to explore whether the publication of grades obtained at a standardized high school test (the *Enem*) resulted in changes in enrollments in high and low performing schools, through a sharp regression discontinuity design. The results show that the disclosure of school grades did not result in students reallocating between both types of school, in neither private nor public schools. The findings remain unchanged when I control for the degree of competition faced by schools or their socio-economic environment.

Keywords: School choice, Standardized tests, School accountability

*I am grateful to my supervisor, Ricardo Madeira, and to Fernanda Estevan and Marcos Rangel for their helpful comments. Email: alepine@usp.br

1 Introduction

Despite significant improvements in access to education, low levels of student learning remain an issue in most developing countries, with international assessments such as PISA[1] showing that many children lag behind regarding basic skills. This has highlighted the need to better measure education quality, and led many governments to collect data on learning outcomes through standardized tests. In addition to guiding educational policies, these indicators are sometimes made public so as to create accountability and help students and their families make informed school choices.

We can typically distinguish between "soft" accountability policies which consist of just reporting information on school performance, and "hard" accountability policies, where there are financial or political sanctions for poor performers. Although soft accountability policies do not involve direct consequences for schools that perform poorly, they are expected to generate pressure on these schools as students can "vote with their feet" and move to a better school.

But whether students and their families value and use the available information on school performance when choosing schools is not clear. There is some evidence that this type of information can affect school choice. Hastings and Weinstein (2008) find that information on school test scores led significantly more parents to choose a better performing school in a natural experiment in North Carolina. Similarly, a study in Canada by Friesen, Javdani, and Woodcock (2009) showed that parents revise their beliefs when information about school quality is provided through report cards, and "vote with their feet" by changing schools. Another branch of the empirical literature that focuses on the relationship between school test scores and house prices also suggests that families react to information on school quality. Using data from the U.S., Black (1999) finds evidence that prices of houses in districts with better schools are higher, and Figlio and Lucas (2004) show that not only school grades, but assignment of schools to discrete letter grades determined by the state have an impact on house prices and residential location decisions[2].

However, other studies show that school performance is not always the main determinant of choice and that preferences regarding schools are heterogeneous across socioeconomic groups. Hastings, Kane and Steiger (2010) estimate a demand model for schools and show that high-income families place a higher weight on school quality, while social

[1] Program for International Student Assessment of the OECD, which evaluates education systems around the world and currently covers 70 countries.

[2] Although school grades might not always be an accurate measure of school quality, as they are also influenced by student composition, they are used by parents as a primary measure of school quality, as pointed by Black (1999).

preferences play a larger role for families from minority or disadvantaged groups. Along the same lines, Elacqua et al. (2006) show that parents value demographics when choosing schools in Chile, and Gallego and Hernando (2009) find that households characteristics influence which schools characteristics are most valued. Interestingly, Gibbons and Machin (2006) find evidence that although parents prefer better performing schools, they also have a preference for "popular" or over-subscribed schools independently of their performance. Finally, Carneiro (2013) shows that in Pakistan parents seem to strongly value distance when choosing schools, while school attributes related to student performance are not valued as much, although this might be because they are not easily observed.

Understanding how much students and their families react to information on education quality is important from a policy perspective. Several countries, including Brazil, have adopted policies consisting on publishing information on school's performance[3] based on the assumption that this will create pressure on schools and that quality is an important criterion for families when choosing schools. However, it is not obvious that this type of policy actually affects students' choices.

In this paper I propose to investigate whether available public information on school quality affects student enrollment choices using administrative data from Brazil. In particular, I want to assess whether the score obtained by schools at a standardized test that covers both private and public high schools, the *Enem*,[4] has an effect on the demand for these schools, as measured by the number of enrolled students. I focus on enrollments in the first year of high school, as it is the start of a new school cycle and many students change schools at that point. To establish causality, I take advantage of an exogenous rule that determined that only schools with a minimum number of test-takers would have their results published. I show that schools do not seem to be able to manipulate the number of test takers, which allows me to use a sharp regression discontinuity design. In order to look at differential effects according to schools' performance, I split the sample between high performing and low performing schools using different criteria.

In addition to student reallocation effects, there can also be a supply-side effect as schools react to increased competitive pressure by improving their quality. Nielson (2013) for example, tries to disentangle both effects in the context of the school voucher policy in Chile. Here I focus exclusively on the demand side, and only look at short term reallocation effects without taking into account possible subsequent changes in school quality.

I find that the sign of the coefficients associated with the discontinuity go in the

[3]In Brazil an indicator of school performance at the national level for basic education (Ideb) is published every two years since 2007, and some states have their own indicators.

[4]The Enem was not originally designed to measure school quality, as will be discussed later.

expected direction in almost all specifications, with the best performing schools showing a positive coefficient and the worse performing schools a negative coefficient. However, the publication of grades at the school level has no significant effect on enrollments for either private or public schools. In baseline estimations, the coefficients point to changes of up to 6% of enrolled students, with the exception of high performing public schools which show coefficients of up to 12%. This finding is robust to the use of different cutoff rules for splitting the sample between high and low performing schools, and does not seem to depend on the degree of competition faced by schools. Similarly, results do not change when taking into account schools' socio-economic environment.

Although there are several studies looking at how accountability policies affect student performance [5], few papers look at how simply disclosing information on school quality affects students' choices and school market outcomes. One example is a study by Mizala and Urquiola (2013), who also use a regression discontinuity design and show that information on school value added did not influence parental choices and school market outcomes in Chile. Additionally, an experiment by Andrabi et al. (2014) in Pakistan shows that the provision of information on test scores had little effects in terms of switching of schools. This paper therefore contributes to this literature by providing new empirical evidence from the Brazilian context.

Brazil offers an interesting case study for two reasons. First, there has been an important effort of collection and dissemination of data on school performance in recent years, which resulted in information on school quality becoming much more accessible. However, this has generated extensive debates over the supposed benefits and disadvantages of the publication of these rankings. As a result, Enem publication rules have been changed twice since the first time test scores were released at the school level in 2006, reflecting a lack of consensus on the ideal policy and pointing to the importance of studying the effects of these policies rigorously. Second, Brazil presents an interesting setting where both private and public schools coexist, and a large performance gap exists between them. However, the extent to which there is competition and migration between both types of school is not well known.

The remainder of this paper is organized as follows. Section 2 provides background on the education system in Brazil and the Enem exam. Section 3 presents the data used and some descriptive statistics. Section 4 details the empirical strategy, and estimation results are reported in Section 5. A few robustness checks are reported in Section 6, and Section 7 presents some concluding remarks.

[5] Most papers study the effects of vouchers or "hard" accountability policies. Koning and van der Wiel (2010) and Camargo et al. (2014) look at the effects of soft accountability policies consisting on the publication of school rankings on student performance.

2 Background: the Education System in Brazil and the Enem Exam

2.1 The Education System in Brazil

The basic education system in Brazil is divided in cycles. After preschool, the second cycle of basic education lasts nine years and is attended by students from 6 to 14 years approximately (primary and middle school), while the third cycle (high school) lasts three years and is attended by students from 15 to 17 years approximately. Both private schools and free public schools coexist. Public schools account for over 85% of enrollments, as shown in Table 1, and can be run by municipal governments, state governments or the federal government[6]. While primary and middle schools are mostly managed by municipal governments, the majority of high schools are managed by state governments.

Access to education has improved in the last fifteen years and is close to that of developed countries - school enrollment was estimated at over 98% for children aged 6 to 14, and at 84% for children aged 15 to 17 in 2012. But although education quality has also improved recently, it remains poor as evidenced by international assessments. Standardized tests also show there is a significant performance gap between public and private schools, with public schools facing poor teacher quality and high teacher absenteeism, as well as high repetition rates and high dropout rates among teenagers.

Private schools can determine their own admission policy and fees, and some schools facing particularly high demand use admission tests, lotteries, or offer places on a "first come first served" basis. Admission to public schools varies from state to state, and some states are more flexible than others concerning school choice. But even in states where admissions are centralized such as the State of São Paulo (the richest and most populous in Brazil), and children are encouraged to attend a school near their home, students can generally apply to other schools and have some degree of choice[7]. In practice, however, there is excess demand for some schools, and admission criteria are not always transparent.

[6] Schools managed by the federal government are much less common and have characteristics that differ from other public schools, such as higher spending per student, better paid and more qualified teachers. Additionally, many offer technical or professional education and have selective entry exams.

[7] In the case of São Paulo for example, families need to provide an address that will be the basis for school allocation, but it is not mandatory that they provide their home address and many provide work or friend's addresses.

2.2 The Enem Exam

The Enem (*Exame Nacional do Ensino Médio*) is a test aimed at students finishing high school in Brazil that happens every year, managed by the Ministry of Education. It is the largest exam in the country, and it is used as part of the selection process of many higher education institutions. Although it was not originally designed to serve as an indicator of school quality, it is generally viewed as such and school rankings based on Enem scores are largely commented and disseminated in the media, as well as used by some private schools as a marketing strategy to attract students. There are other indicators that were explicitly designed to measure school quality in Brazil, such as the Ideb (*Índice de Desenvolvimento da Educação Básica*), or specific tests created by state governments. However, the Enem is so far the only exam that covers both public and private schools at the national level, and therefore allows students and their families to compare the performance of a large number of schools. It is also the only exam at the national level that is available yearly. According to official statistics, the percentage of eligible schools participating in the exam was close to 80% in 2006, and more recent data suggest this percentage has increased since.

Although the exam is not mandatory, an increasing number of students take it each year: the number of registered students has gone up from around 150,000 in 1998, its first year, to over 7 million in 2013. This can be explained to a large extent by the fact that exam stakes have increased in recent years. The Enem has been used as one of the criteria for government scholarship attributions for disadvantaged students who want to pursue higher education since 2005 (through the *Prouni* program), and many higher education institutions started using Enem scores as the only entry requirement through a unified selection process. In addition, it also now serves as the equivalent of the high school diploma for students over 18 years old who have not completed high school. The vast majority of students who take the test are in the last year of high school, but younger high school students also take the test for training, as well as individuals out of the school system or in special education schemes. However, school grades are computed taking into account only students enrolled in the last year of high school.

Until 2008, the Enem included one essay and 63 multiple choice questions and grades were given on a scale from 0 to 100. From 2009 onwards, the test format changed considerably: the number of multiple choice questions went up to 180 and included more subjects, grades were given in a scale from 0 to 1000, and test scores were calculated using Item Response Theory[8]. Enem scores at the school level started being publicly released from

[8]According to this methodology, the probability of obtaining a correct answer is assessed according to its difficulty, the probability that a student could guess a correct answer, and its ability to discriminate against students. As a consequence, test scores only started being comparable

2006, for grades obtained in the previous year. The exam usually happens in the second semester of the Brazilian school year (between August and December), and individual grades are released a couple of months afterwards. Grades at the school level are then later publicly released on the internet website of Inep, a government body related to the Ministry of Education, and largely commented by newspapers and magazines. Figure 1 shows a few examples of published rankings, and Figure 2 presents the evolution of internet searches on Enem rankings using Google trends, showing increases in searches just after school grades are released. The release calendar of Enem test scores since the first publication is presented in Table 2.

Since the first time grades were disclosed at the school level, a rule stated that only schools with 10 or more test takers enrolled in the final year of high school would have their grades released. This was related to concerns that grades of schools with fewer test-takers might not be representative due to student selection. In subsequent years, additional criteria for publicly releasing school grades were gradually introduced: in 2010 (relative to the Enem 2009) it was decided that only schools where test-takers represented at least 2% of total enrollment would have their grades released, and from 2012 this percentage was raised to 50% of total enrollment. These changes, in addition to modifications to the format of school rankings, were an attempt to avoid misleading comparisons, and a response to criticism that followed the publication of previous school rankings. Despite these efforts, school rankings based on Enem scores remain widely disseminated in the media.

3 Data

3.1 Enem Data

This paper uses Enem microdata between 2005 (the first session for which grades were publicly released at the school level) and 2008. I do not use data from 2009 onwards, as the exam format and publication rules have changed from that year, as described previously. It is not clear whether students and their families continued to interpret the scores the same way as before, and more importantly, the new requirement that at least 2% of enrolled students participated in the exam for a school to have its grades published means the same regression discontinuity design cannot be applied.

The Enem data provides information on each student that has taken the exam

over time from 2009 onwards. In previous years it was only possible to compare different schools' scores in the same year, as the difficulty of the test varied each year.

including grades, socio-economic background, and school attended. I am therefore able to calculate average test scores for all schools, including those that did not have their grades released because less than 10 students took the exam. Each student has an essay grade, as well as a test grade from the score obtained in multiple choice questions, and each school's total grade is calculated as the average of these individual grades. Only students enrolled in the final year of high school, and who were actually present the day of the exam are included in the calculation. I exclude federal schools, as they are governed by specific rules and represent less than 1% of the sample, as well as schools that are temporarily or permanently closed. By calculating the number of eligible students per school, I am able to create a dummy variable indicating whether schools had their scores published or not.

3.2 School Census Data

Schools from the sample are then matched to school census data through a unique school identifier. This allows me to obtain information on the total number of students enrolled in each grade in schools from the sample, as well as calculate the percentage of students enrolled in the final year of high school who took the Enem Exam. In order to consider the necessary delay for Enem results to affect students' enrollment decisions, I look at enrollment data two years after the exam takes place (that is, for the Enem session that takes place at t=0, I look at enrollments in t=2). This is because Enem scores at the school level are only published the year following the exam, at a time when enrollment decisions for the year in question have already been made. Therefore, any reallocation effects could only be observed two years after the exam takes place[9]. Over 97% of schools from the original sample could be matched to school census enrollment data two years later.

Some schools in the school census database report having zero students enrolled in one or more high school grades. Although it is possible that some small schools do not have any students enrolled in a particular grade, most of these cases are in all likelihood missing data. Most schools that report having no students in a specific grade also report having no students in all the other grades. Moreover, there is a significant discontinuity in the frequency of schools that report having zero students enrolled in a given grade and schools that report having one student enrolled, while no such drops are seen at other points. For this reason, I treat this data as missing, which represent approximately 10% of the sample[10].

[9]The school year in Brazil starts in February, but enrollment decisions for students are typically made earlier, between October and December of the previous year. School census data are collected in May each year

[10]Although these schools have slightly lower Enem averages than the rest of the sample, the

It is more common for students to change schools between school cycles, for example between the last year of middle school and the first year of high school, than in other grades. This is the case partly because some schools only offer a specific cycle of education, and therefore some students are obliged to change schools if they want to continue studying. I therefore focus on enrollments in the first year of high school as the outcome variable, as any potential reallocation effects are likely to be stronger at that moment.

3.3 Descriptive Statistics

Table 3 provides summary statistics on the final sample of matched schools, which includes a total of 91,457 schools across 5,240 municipalities when pooling together all four years of data. There are considerable differences between private and public schools in terms of student achievement at the Enem test, with private schools scoring between 25% and 30% higher on the test in the period considered. Private schools are also much smaller on average - less than half the size of public schools - and are attended by students with a more privileged socio-economic background.

The percentage of eligible students taking the Enem test, calculated as the ratio between total Enem takers and total enrolled students in the final year of high school, has increased with time as the exam gained in importance and involved higher stakes for students. In private schools the ratio increased from 48% to 66%, and in public schools it increased from 42% to 49% between 2005 and 2008. As expected, private schools show a higher percentage of students taking the test, as private school students have a higher probability of pursuing higher education.

The 2005 and 2006 School Census databases have a different format and a much higher number of schools with missing enrollment data than the 2007 and 2008 School Censuses. In particular, small schools are much more likely to have missing data than bigger schools [11]. As a result, we can see sizeable differences in the number of enrolled students from 2007 onwards compared to 2005 and 2006. Although this could result in the sample composition of schools being different across years, it is unlikely to affect the analysis since I compare schools from both sides of the discontinuity inside a window of data, and no significant differences in school characteristics were found bewteen both sides, as will be shown in the next section.

The analysis in this paper focuses on a window of data around the discontinuity

grades do not differ significantly between schools from each side of the discontinuity in the window of data considered in the analysis.

[11] This can be inferred by looking at enrollments in 2007 and 2008 for the subset of schools with missing data in 2005 and 2006

of 5-15 Enem takers, and therefore concerns a specific subset of schools with a small number of test takers. Table 4 provides a sense of how these schools differ from others, and shows that there is a higher proportion of private schools among them, and that they are generally smaller. They also have a smaller fraction of students who take the exam (around 22% of students take the exam in schools with up to 5 Enem takers, compared to over 50% for schools with more than 25 Enem takers). This might mean there is a higher selection of students who take the exam, although average grades are not very different from grades of larger schools.

4 Empirical Strategy

4.1 Methodology

The question this paper wants to address is whether schools' performance in the Enem exam has an effect on students' school choice, and therefore on enrollments in the first year of high school. If families take into account quality when choosing a school for their children, then we would expect schools that had a low grade published to attract less students than similar schools that did not have their grades published. Similarly, we would expect schools with relatively good grades to attract more students if their results are published.

To isolate the effect of disclosing grades at the school level, I take advantage of the discontinuity created by the rule that sets the 10 student threshold for publishing Enem results. Additionally I split the sample in two, between high performing schools and low performing schools, in order to look for different effects according to the type of school. As the criteria used for splitting the sample is arbitrary, I use a series of different cutoff rules to assess the robustness of the results.

The discontinuity created by the 10 student threshold creates a quasi-experimental setting that allows me to estimate the effect of publishing grades at the school level by using a sharp regression discontinuity design. The fact that an exogenous rule determines which schools will have their grades published means that if schools are unable to precisely manipulate that rule, a possibility that will be discussed in the next subsection, then those just above the cutoff (the "treated" schools) can be considered a good counterfactual to those just below the cutoff (the "control" schools), and by restricting our attention to data close enough to the discontinuity, we are in a similar case as a local randomized experiment.

The fundamental hypothesis that allows identification in this case is that the conditional expectation of outcome Y_i (the number of enrolled students) with respect to the assignment variable X_i (the number of Enem takers), is continuous at the cutoff point c. This smoothness assumption is necessary because we only observe individuals from one or the other side of the cutoff and never both at the same time. Using the potential outcomes framework and following the notation on Lee and Lemieux (2010) and Imbens and Lemieux (2007) , if $Y_i(1)$ is the outcome for treated schools and $Y_i(0)$ is the outcome for control schools, we want to estimate:

$$\lim_{\epsilon \downarrow 0} E[Y_i(1)|X = c + \epsilon] - \lim_{\epsilon \uparrow 0} E[Y_i(0)|X = c + \epsilon]$$

which is equal to:

$$E[Y_i(1) - Y_i(0)|X = c]$$

If $E[Y_i(1)|X]$ and $E[Y_i(0)|X]$ are continuous at the cutoff point c, then any discontinuity of the conditional function at the cutoff can be attributed to the effect of the treatment.

To estimate the effect of the publication of grades at the school level, I run OLS regressions separately on both sides of the discontinuity, which is the equivalent of estimating the following equation:

$$log(Y_{it+2}) = \beta_0 + \phi_1(X_{it} - c) + \beta_1 D_{it} + D_{it}\phi_2(X_{it} - c) + \epsilon_{it} \tag{1}$$

Where Y_{it+2} is the outcome variable (the number of enrolled students in the first year of high school in school i, in year $t + 2$), D_{it} is a dummy variable equal to 1 if school i had its grades published in year t, X_{it} is the assignment variable (the number of Enem takers), $\phi_1(.)$ and $\phi_2(.)$ are polynomials and c is the cutoff point which equals 10. Taking the log of the outcome variable allows me to approximate the percentage change in enrollments and deal with outliers. Since private schools are considerably smaller than public schools on average, it also facilitates the comparison of results. Although there are several years of data, I do not use fixed effects at the school level. The reason is that if some schools stay on the same side of the discontinuity across the years, then school fixed effects would capture the effect of the publication of grades, and these schools would effectively be excluded from the analysis.

4.2 Internal Validity

The regression discontinuity design might be invalidated if schools were able to manipulate the number of students who take the exam (for example encouraging or discouraging students to take the Enem), and influence whether or not their results are published. In this case assignment to either side of the discontinuity would not be random and could be correlated with schools' characteristics. If for example bad schools discouraged their students from taking the exam in order to avoid having a low grade published and preserve their reputation, but for some reason only small schools succeeded in doing so, then there would be a higher proportion of larger schools at the right side of the discontinuity. In this case one could erroneously conclude that the publication of test scores increases enrollment for bad schools, when this result is just driven by a change in school composition across each side of the discontinuity.

In practice, however, it is unlikely that schools are able to manipulate the number of Enem test takers. First, policies influencing interest for the Enem, such as the attribution of scholarships based on Enem grades are determined at the federal level and cannot be directly influenced by individual schools. Second, it is the number of actual test takers that is taken into account by the publication rule, and not the number of students who enrolled for the test. This means that students who are absent the day of the test are not counted, making it more difficult for schools to influence the number of test takers.

A more formal test to verify this and establish the internal validity of the regression discontinuity methodology is to look at jumps in the density of schools around the threshold, following the method proposed by McCrary (2008). A density plot of schools, as presented in Figure 3, does not suggest there are any jumps. Although there is a high frequency of schools with only one student taking the Enem, this will not be a concern for the analysis since I focus on a window of data closer to the discontinuity (5-15 students in baseline specifications). Similarly, a density smoothness test obtained by estimating a local polynomial on both sides of the discontinuity, shown in Figure 4, does not suggest any jumps around the threshold. This result is true for different choices of bandwidth and polynomial degrees, and also holds when looking separately at private and public schools (not shown here).

Another way of assessing whether there might any form of manipulation is to look for jumps in covariates around the discontinuity. If the only difference between treated and control schools at the cutoff area is the assignment rule, then there would be no reason to see jumps in observable school characteristics. As an additional test of the validity of the methodology, I run a series of regressions using a similar specification as in equation (1) where the explanatory variable is a set of covariates at the school level in t, in order to see whether the dummy coefficient β_1 is statistically significant. I only include observations from a symmetrical window of 5-15 Enem takers around the discontinuity.

I look at possible jumps in grades, socio-economic variables that could be a proxy for student composition, and in enrollment pre-treatment data. The results presented in Table 5 generally suggest there are no jumps in covariates, but the dummy coefficient is significant for enrollment pre-treatment data when I add a quadratic term. Further investigation shows this result is not robust to minor specification changes such as adding a cubic term, or when I split the sample by school type and school quality (not shown here). Although this suggests it is unlikely that this jump is driven by differences in school composition around the discontinuity that could affect the outcome of interest, I control for enrollment in t in all the specifications to deal with any possible confounding factors. I also run the same regressions separately for private and public schools for all the other covariates (shown in the Appendix), and conclusions remain the same.

Additional evidence on the randomness of schools' position around the threshold can be found by looking at dynamics. If a school is able to manipulate the number of students taking the exam, then it is likely that some schools will systematically fall on the same side of the threshold in different years. Table 6 shows, for a given year, the proportion of schools that stay on the same side of the discontinuity the following year. The first two columns show that this proportion is very close to the proportion of schools that falls on the other side of the discontinuity, suggesting schools are not able to precisely control their position.

5 Results

5.1 Different Cutoffs for High and Low Performing Schools

I first look at how enrollments in high and low performing schools react to the publication of Enem grades at the school level. Given the important institutional differences between public and private schools in Brazil, and as private schools have much more flexibility in their admissions procedures, I run separate regressions for each type of school.

I estimate equation (1), where I control for enrollment in $t = 0$ and include year dummies. Standard errors are clustered at the level of municipalities. Adding controls helps reduce sampling variability, but should not change the overall results. In this case, controlling for matriculation levels in $t = 0$ is particularly important since I found jumps in pre-enrollment data in some specifications at the cutoff. The reason for including year dummies is that the release date of Enem results has changed slightly across years, which could affect the degree to which students react to the publication of grades.

Regression discontinuity analysis usually implies a tradeoff between the number of observations that can be used in the analysis, and the size of the potential bias in the estimated results. The narrower the window of data used, the smaller the sample size. But this also decreases the probability of including in the analysis schools that are too different from each other, and therefore of having unobservable factors correlated with the outcome variable driving the results. I therefore focus on a narrow symmetrical window of observations across both sides of the discontinuity, of 5-15 students, and present robustness checks with slightly larger windows in the Appendix.

In order to separate high and low performing schools, I first divide the sample in two using the 50 mark in a scale of 0-100 as a cutoff, which could be interpreted as a "psychological" threshold. Results are presented in column (1) of Table 7. The publication of school grades does not lead to a significant change in the number of enrolled students in either good or bad schools. However, the sign of the dummy coefficient goes in the expected direction in most cases, with good schools receiving more students and bad schools loosing students relative to similar schools that did not have their grades published. As the dependent variable is in log, the coefficients should be interpreted as percentage changes. In most cases the effect is very small, but interestingly, the best performing public schools show the higher gain, of 11% to 12% additional students, although not statistically significant.

It is possible, however, that absolute grades are not the relevant metric used for comparing schools and that families consider private and public schools as separate markets. In fact, the media sometimes presents separate rankings by school category. If that is indeed how comparisons are made, using a fixed cutoff for both types of school might be misleading. As illustrated in Figure 5, there is a considerable achievement gap between private and public schools which means the 50 cutoff includes only the 15% best public schools but includes the 80% best private schools, which might help explain the lack of effects for private schools.

To allow for this possibility, I create separate rankings for each type of school. Columns (2) and (3) of Table 7 present results where the sample includes only the 40% and 20% better and worse schools of each category respectively. Results are similar as before, with coefficient signs going in the expected direction but no significant effect from the publication of grades on enrollment. Choosing the 40% best performing public schools is equivalent to using a cutoff grade of 45 out of 100, which, explains the lower coefficients obtained for these schools in column (2) as compared with column (1). In column (3), the coefficients for these schools are very close to the first set of results, which use a cutoff of

50, as this is equivalent to using a cutoff of 48.

The fact that the sample size is relatively small for some subgroups means I may have low statistical power to detect any effects. For a more systematic analysis and to see whether there is a general pattern as I progressively restrict the cutoffs for high and low performing schools, I run estimations using different percentiles and plot the coefficients obtained in a graph (Figure 6). Results do not suggest a clear pattern, although high performing public schools have more consistently positive coefficients.

5.2 Degree of Competition Faced by Schools

The competitive environment faced by each school can vary greatly, with some schools facing strong competition and others facing no competition at all, such as isolated schools in rural areas. The average number of high schools by municipality is around 9 in the sample used in the estimations (that is, considering the window of data of 5-15 Enem takers) but there is great variation, with some municipalities having just one school and the largest, São Paulo, with close to 1200 schools.

Therefore, results from previous estimations suggesting that the publication of grades at the school level did not affect enrollment could be masking significant disparities, with reallocation effects only taking place in municipalities where schools face a more competitive environment.

In order to take this into account, I create two different measures of school competition. The first is the Herfindahl-Hirschman Index (HHI), a commonly used measure of market concentration, which I adapt to the case of schools. The HHI is usually calculated as the sum of squares of market shares of firms within an industry, and can range from close to 0 (in the case of a very competitive market) to 1 (in the case of a single monopoly). To adapt the HHI to the case of schools, I calculate for each school the share of students enrolled in high school as a percentage of total high school students in the municipality, according to the formula:

$$HHI_{it} = \sum_{i=1}^{N} share_{it}^{2}$$

Where $share_{it}$ is the market share of school i in year t. For practical purposes, I make abstraction of the fact that students can attend schools in adjacent municipalities and consider a municipality as a school market.

The second measure I use is the share of high school students enrolled in private schools in each municipality, which is another measure of school competition commonly found in the literature (Hoxby, 1994) and can be considered a proxy for the degree of pressure faced by public schools.

To take into account the degree of competition faced by schools in previous estimations, I include the interaction term $Comp_{it} * D_{it}$ in the baseline specification (which uses the 50 grade cutoff), where $Comp_{it}$ is each one of the competition measures mentioned before. I also control for the level of competition, as in equation (2):

$$log(Y_{it+2}) = \beta_0 + \phi_1(X_{it} - c) + \beta_1 D_{it} + D_{it}\phi_2(X_{it} - c) + \beta_2 Comp_{it} + \beta_3 D_{it} Comp_{it} + \epsilon_{it} \quad (2)$$

The inclusion of interacted terms changes the interpretation of the coefficients. When the interaction term is included, β_1 measures the effect of the dummy on enrollment when $Comp_{it}$ is 0. To facilitate the interpretation of the coefficients, I center the competition variable so that its average is 0 and β_1 measures the effect of the dummy on enrollment when $Comp_{it}$ is at its average value.

Results are presented in columns (1) and (2) of Table 8. The higher the Herfindahl index, the lower the level of competition among schools in a given municipality. Therefore, if competition among schools affects the intensity of reallocation effects, we could expect the interaction term to be positive for bad schools (which would lose fewer students when there is little competition), and negative for good schools. The dummy coefficients are still not significant and neither are the interaction terms, suggesting that increased school competition does not lead to stronger reallocation effects in response to the publication of school grades.

To take into consideration the fact that competition does not only operate in terms of the quantity of schools available but also depends on the variety of schools in terms of quality, I create an index to measure the dispersion of school grades in each municipality using the same logic of the Herfindahl-Hirschman Index presented above. Instead of calculating market shares, I create 10 artificial grade categories of 10 point intervals (0-10, 10-20 etc.) and then calculate the total number of schools that fall in each category. I then calculate the share of schools represented by each category as a percentage of total schools in the municipality. Results using this measure of competition are presented in column (3) of Table 8 and do not alter previous conclusions.

5.3 Schools' Socio-Economic Environment

I next consider the socio-economic environment of schools at the level of municipalities. Different studies have showed that school preferences are heterogeneous among socio-economic groups, and in particular the degree to which school quality is valued. School choice decisions could therefore be correlated with socio-economic variables, such as education or the income level. To account for this, I interact the treatment dummy with income per capita and the average number of years of education in each school's municipality. I also run regressions where I interact the treatment dummy with income inequality at the municipality level. The level of income inequality in the municipality could also have an impact on reallocation effects, as in places with high income inequality schools are likely to have more scope for reacting to changes in demand by adjusting prices.

Socio-economic data for municipalities is available from IPEA (*Instituto de Pesquisa Econômica Aplicada*, a Brazilian government think thank), at a decennial frequency. Inequality data is obtained from the *Atlas do Desenvolvimento Humano no Brasil* [12]. I use data from the year 2000 for all indicators considered.

As in previous estimations, baseline regressions are run with a specification similar to equation (2), where the treatment dummy is interacted with socioeconomic variables at the municipality level, designed by SE_{it}. Estimation results are presented in table 9, and do not suggest any of the variables considered has an influence on student reallocation effects. The dummy coefficients remain small in most cases, with the exception of the sample of high performing public schools, although they are generally not significant.

6 Robustness Checks

I next present some robustness checks to test the validity of the results obtained. Typical threats to regression discontinuity designs, such as the self-selection of schools around the discontinuity and jumps in covariates have been addressed previously, so these can be seen as complements to previous tests. I do not present robustness tests using higher order polynomials, as these are not recommended in regression discontinuity analysis (Gelman and Imbens, 2014).

First, baseline results are replicated using slightly larger windows of data, of 4-16 Enem takers and 3-17 Enem takers. Results are presented in the Appendix and are very similar to those obtained previously. Second, I run the same regressions using as outcome

[12] Atlas of Human Development in Brazil, produced in conjunction by UNDP Brazil, IPEA and the João Pinheiro Foundation. http://www.atlasbrasil.org.br/2013/pt/

variable the total number of enrolled students in high school, and not only in the first year of high school. Although migration effects are expected to be lower in higher grades, this allows me to take into account possible dropout effects in these higher grades. Although the magnitudes and signs of the coefficients obtained are more variable in this case, they remain non-significant.

As a last exercise I create local school rankings inside each of the 26 states and federal district of Brazil, account for the possibility that the relevant comparison between schools is made locally. For each state, I divide the sample between high and low performing schools by taking the 50% best and worst schools of each state respectively. Although some samples are small and therefore the estimates obtained should be interpreted with caution, previous conclusions are unchanged.

7 Conclusion

School accountability policies have been at the center of debates on how to improve education quality in both developing and developed countries. In Brazil, as in countries facing similar issues, there has been controversy about the effectiveness of soft accountability policies consisting of reporting information on school quality as a way to pressure schools.

I take advantage of a discontinuity in the rule concerning the publication of school grades of a major high school exam, the Enem, to look at short term student reallocation effects between high performing schools and low performing schools. Although not specifically designed as a tool for school accountability, the Enem is seen in practice as a measure of school quality and school rankings based on Enem scores are widely commented in the media.

Despite the attention drawn by these rankings, I do not find any significant changes in enrollment in either private or public schools. This finding is unchanged when the treatment effect is interacted with different measures of school competition or with socio-economic variables at the level of municipalities. Tests of internal validity and robustness checks confirm the validity of the results obtained. These findings are in line with Mizala and Urquiola (2013), who find no effect of disclosing information on schools' value added in Chile.

A series of explanations can be put forward to explain these results. First, good private schools might be capacity constrained and thus prefer to adjust prices or select students based on ability rather than accept more students. Unfortunately data on private school fees at the national level are not available, but this is a possible outcome that should be explored further. However, this would not explain why bad performing schools do not loose students. Another possible explanation is that the information does not effectively reach parents. Grades at the school level are originally published on-line, which limits its reach to families with access and knowledge on how to use the internet. And although rankings are commented in the media, they are sometimes restricted to the 1000 or 100 top schools, or restricted to schools in a given state, which might not help all families make informed decisions. Finally, school performance might not be the main criterion of choice for the majority of families, and other factors such as distance or educational philosophy might be privileged. This would be in line with studies pointing to heterogeneous preferences regarding school quality. It might be that only a very small fraction of the population cares about high school rankings, the most privileged and whose children want to pursue higher education, which might explain why no significant effects were found even when accounting for municipalities socioeconomic variables in the regressions.

An important point to consider is the external validity of the findings. The regression discontinuity analysis produces local average treatment effects, which apply to the subpopulation of schools studied, and might not be generalizable to larger schools. With this caveat in mind, the findings seem to suggest that simply disclosing information might not be sufficient to generate significant student reallocation effects and influence families to exert school choice. However, further analysis is necessary to understand the conditions under which soft accountability policies can be effective, as it is likely that the effects of this type of policy will be very context-dependent, with factors such as how the information is disseminated, local preferences regarding school quality, and degree of school choice playing an important role.

References

Andrabi, T., Das, J., and Khwaja, A. I. (2014). Report Cards : The Impact of Providing Test-score information on Educational Markets. *HKS Working Paper No. RWP14-052*, (June).

Black, S. E. (1999). Do Better Schools Matter ? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics*, 114(2):577–599.

Camargo, B., Camelo, R., Firpo, S., and Ponczek, V. (2014). Information, Market Incentives, and Student Performance. *IZA Discussion Paper*, (7941).

Carneiro, P., Das, J., and Reis, H. (2013). Parental Valuation of School Attributes in Developing Countries: Evidence from Pakistan. *mimeo, University College London.*

Elacqua, G., Schneider, M., and Buckley, J. (2007). School Choice in Chile: Is It Class or the Classroom? *Journal of Policy Analysis and Management*, 25(3):577–601.

Figlio, D. N. and Lucas, M. E. (2004). What's in a grade? School report cards and the housing market. *American Economic Review*, 94(3):591–604.

Friesen, J., Javdani, M., Smith, J., and Woodcock, S. (2012). How do school 'report cards' affect school choice decisions? *Canadian Journal of Economics*, 45(2):784–807.

Gallego, F. and Hernando, A. (2010). School choice in Chile: Looking at the demand side. *Pontificia Universidad Catolica de Chile, Documento de Trabajo, (356).*

Gelman, A. and Imbens, G. (2014). Why High-order Polynomials Should not be Used in Regression Discontinuity Designs. *National Bureau of Economic Research Working Paper Series*, No. 20405.

Gibbons, S. and Machin, S. (2006). Paying for Primary Schools : Admission Constraints, School Popularity or Congestion? *The Economic Journal*, 116(510).

Hastings, J. S., Kane, T. J., and Staiger, D. O. (2008). Heterogeneous Preferences and the Efficacy of Public School Choice. *NBER Working Paper, 2145.*, (May).

Hastings, J. S. and Weinstein, J. M. (2008). Information, School choice and Academic Achievement: Evidence from Two Experiments. *Quarterly Journal of Economics*, 123(4):1373–1414.

Hoxby, C. (1994). Do private schools provide competition for public schools? *NBER Working Paper No 4978.*

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.

Koning, P. and van der Wiel, K. (2010). School responsiveness to quality ranking: An empirical analysis of secondary education in the Netherlands. *IZA Discussion Paper*, (4969).

Lee, D. S. and Lemieux, T. (2010). Regression Discontnuity Designs in Economics. *Journal of Economic Literature*, 48(June):281–355.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.

Mizala, A. and Urquiola, M. (2013). School markets: The impact of information approximating schools' effectiveness. *Journal of Development Economics*, 103:313–335.

Neilson, C. (2013). Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students. *Job Market Paper.*

Table 1: Enrollment by type of school in basic education (2012 School Census)

	% total enrollment	
	Primary and middle school (ages 6 to 14)	High school (ages 15 to 17)
Private schools	14.3%	12.7%
Public schools	85.7%	87.3%
State schools	30.6%	84.9%
Municipal schools	55%	0.9%
Federal schools	0.1%	1.5%

Table 2: Enem results release calendar

Enem session	Exam date	Individual grades release	School grades release
Enem 2005	September 2005	November 2005	February 2006
Enem 2006	August 2006	November 2006	February 2007
Enem 2007	August 2007	November 2007	April 2008
Enem 2008	August 2008	November 2008	April 2009
Enem 2009	December 2009*	January 2010	July 2010
Enem 2010	November 2010	January 2011	September 2011
Enem 2011	October 2011	December 2011	November 2012
Enem 2012	November 2012	December 2013	November 2013
Enem 2013	October 2013	January 2014	December 2014

* The 2009 exam was delayed as there were fraud suspicions

Table 3: Summary statistics

	2005		2006		2007		2008	
	private	public	private	public	private	public	private	public
Panel A School characteristics								
No. of schools	6091	15946	5841	15439	5846	17259	6812	18223
No. of schools with < 10 Enem takers	1764	1891	1694	1831	1898	2604	2071	2704
Avg. grade (out of 100)	55.3	42.1	51.9	41.4	61.7	47.9	57.1	45.2
Avg. enrollment in 1^{st} year of high school	101	248	100	234	51	182	51	173
Avg. enrollment in 2^{nd} year of high school	103.6	191.8	99.5	180.4	47.4	138.1	46.6	130.5
Avg. enrollment in 3^{rd} year of high school	123.4	162.5	112.4	154.7	46.3	115.6	45.7	111.5
% eligible students taking Enem	47.6	42	51.4	44.3	60.6	46	66	49.4
% of schools which also offer primary/middle education	84	84	83	84	83	83	83	83
Panel B Characteristics of Enem takers								
% Black	5	9	5	9	5	9	5	9
% Low income*	14	53	17	57	16	57	17	55
% of which father has finished high school	55	13	57	14	61	15	58	15
% of which mother has finished high school	62	18	64	19	68	21	65	21
Avg. age	17.2	18.8	18	20.6	17.9	20.6	18	19.6

*Includes the two lower categories of revenue among seven categories in the survey, equivalent to up to 2 minimum wages between 2005 and 2008 (600 to 830 BRL or 200 to 270 USD approximately)

Table 4: Summary statistics by windows of Enem takers

No. Enem takers	1-5	6-10	11-15	16-20	21-25	26+
Avg. grade (out of 100)	47.3	47.6	48.1	48	47.7	47.4
% Low income*	41	42	43	44	46	47
% Black	7	7	7	7	8	8
Avg. enrollment in 1^{st} year of high school (t0)	63	60	72	87	104	233
% of eligible students taking Enem	21.8	39.5	45.8	47.8	48.4	53.1
% Private schools	45	44	40	36	30	16
% Rural schools	13	11	8	6	5	1
Total obs.	9131	9174	8729	7861	6738	49824

*Includes the two lower categories of revenue among seven categories in the survey, equivalent to up to 2 minimum wages between 2005 and 2008 (600 to 830 BRL or 200 to 270 USD approximately)

Table 5: Jumps in covariates estimation

	Avg. Grade		% low income*		% black		% father w/ high school	
Dummy coef.	-0.291	0.536	0.006	-0.010	0.005	-0.002	-0.001	0.020
(Enem takers>=10)	(0.24)	(0.43)	(0.01)	(0.02)	(0.00)	(0.00)	(0.01)	(0.01)
Quadratic	No	Yes	No	Yes	No	Yes	No	Yes
Avg. at left of discontinuity	47.5		0.42		0.07		0.28	
No. obs.	19635		19635		19635		19636	

	% mother w/ high school		Log enrollment 1st year of high school		Log enrollment high school (total)	
Dummy coef.	0.001	0.019	0.009	-0.122**	0.008	-0.125**
(Enem takers>=10)	(0.01)	(0.02)	(0.03)	(0.06)	(0.03)	(0.05)
Quadratic	No	Yes	No	Yes	No	Yes
Avg. at left of discontinuity	0.35		3.7		4.5	
No. obs.	19636		11031		11196	

Controls are year dummies. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations.
*Includes the two lower categories of revenue among seven categories in the survey, equivalent to up to 2 minimum wages between 2005 and 2008 (600 to 830 BRL or 200 to 270 USD approximately)
Standard errors in parentheses
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 6: Dynamics of schools' position around the discontinuity(window of data of 5-15 Enem takers)

	Enem takers >=10 in $t+1$	Enem takers <10 in $t+1$	Outside window	Not found in $t+1$	Total
$\boldsymbol{t}$ **= 2005**					
Enem takers >=10	21.6%	17.8%	47.1%	13.5%	100%
Enem takers <10	21.3%	25%	35.2%	18.6%	100%
$\boldsymbol{t}$ **= 2006**					
Enem takers >=10	23.2%	22.5%	43.2%	11.1%	100%
Enem takers <10	18%	25%	38%	19%	100%
$\boldsymbol{t}$ **= 2007**					
Enem takers >=10	23%	17%	55.2%	4.7%	100%
Enem takers <10	21.6%	26.5%	41.9%	10.1%	100%

This table shows for a given year (t), the proportion of schools that stay on the same side of the discontinuity the following year (t+1)

Table 7: Estimates for different grade cutoffs

	50 mark		40% top/bottom schools		20% top/bottom schools	
Best performing schools	(1)		(2)		(3)	
Treatment effect - Private	0.004	0.025	0.01	0.058	0.063	-0.043
	(0.04)	(0.07)	(0.05)	(0.10)	(0.08)	(0.17)
No. obs	3194	3194	1706	1706	789	789
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	0.119	0.111	0.028	0.021	0.116*	0.118
	(0.08)	(0.17)	(0.04)	(0.09)	(0.06)	(0.12)
No. obs	873	873	2493	2493	1305	1305
Quadratic	No	Yes	No	Yes	No	Yes
Worst performing schools						
Treatment effect - Private	0.003	-0.063	-0.049	-0.056	0.017	-0.001
	(0.12)	(0.20)	(0.07)	(0.12)	(0.12)	(0.20)
No. obs	373	373	1015	1015	369	369
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	-0.015	-0.043	-0.008	-0.003	-0.018	0.032
	(0.03)	(0.06)	(0.04)	(0.07)	(0.05)	(0.10)
No. obs	5774	5774	2969	2969	1715	1715
Quadratic	No	Yes	No	Yes	No	Yes

The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in t=0. In column (1) the best and worst schools are separated using the 50 mark cutoff. In column (2) a ranking of schools is made inside each category (private or public) and the best and worst schools are the 40% higher and lower performing schools. In column (3) the best and worst schools are the 20% higher and lower performing schools inside each category. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations.
Standard errors in parentheses
$^{*}p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01$

Table 8: Estimates including measures of competition (cutoff: 50 mark in 0-100 scale)

	HHI schools (1)		% private schools (2)		HHI grades (3)	
Best performing schools						
Treatment effect - Private	-0.006	0.016	0.037	0.077	0.040	0.080
	(0.04)	(0.07)	(0.04)	(0.07)	(0.04)	(0.07)
Interaction term	0.005	0.005	-0.610	-0.618	-0.535**	-0.542**
$Comp_{it} * D_{it}$	(0.07)	(0.07)	(0.27)	(0.28)	(0.23)	(0.23)
No. obs	3194	3194	3194	3194	3194	3194
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	0.127*	0.128	0.128	0.129	0.127*	0.131
	(0.08)	(0.17)	(0.08)	(0.17)	(0.08)	(0.17)
Interaction term	-0.040	-0.042	0.059	0.070	0.078	0.090
$Comp_{it} * D_{it}$	(0.09)	(0.09)	(0.34)	(0.35)	(0.29)	(0.30)
No. obs	873	873	873	873	873	873
Quadratic	No	Yes	No	Yes	No	Yes
Worst performing schools						
Treatment effect - Private	-0.022	-0.096	-0.009	-0.129	0.027	-0.082
	(0.12)	(0.20)	(0.13)	(0.21)	(0.13)	(0.20)
Interaction term	-0.124	-0.143	0.65	0.695	-0.298	-0.296
$Comp_{it} * D_{it}$	(0.22)	(0.22)	(0.80)	(0.81)	(0.67)	(0.69)
No. obs	373	373	372	372	373	373
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	-0.012	-0.042	-0.011	-0.038	-0.012	-0.039
	(0.03)	(0.06)	(0.03)	(0.06)	(0.03)	(0.06)
Interaction term	-0.013	-0.012	0.317	0.314	0.215	0.213
$Comp_{it} * D_{it}$	(0.04)	(0.04)	(0.25)	(0.25)	(0.20)	(0.20)
No. obs	5774	5774	5774	5774	5774	5774
Quadratic	No	Yes	No	Yes	No	Yes

The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and pre enrollment data. All regressions use the baseline specification where the 50 mark cutoff is used. In column (1) the measure of competition controlled for is the HHI of school market shares. In column (2), the competition variable used is the share of high school students enrolled in private schools in each municipality. In column (3) the competition variable used is the HHI of the dispersion of school grades. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations.
Standard errors in parentheses
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Table 9: Estimates including socio-economic factors (cutoff: 50 mark in 0-100 scale)

	Gini coefficient		Income per capita		Years of education	
Best performing schools	(1)		(2)		(3)	
Treatment effect - Private	0.005	0.026	0.002	0.029	0.001	0.04
	(0.04)	(0.07)	(0.04)	(0.07)	(0.04)	(0.07)
Interaction term	-0.003	-0.003	-0.000	-0.000	-0.006	-0.006
$SE_{it} * D_{it}$	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
No. obs	3194	3194	3190	3190	3190	3190
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	0.128	0.119	0.131*	0.154	0.132*	0.149
	(0.08)	(0.17)	(0.08)	(0.17)	(0.08)	(0.17)
Interaction term	0.003	0.004	-0.000	-0.000	-0.006	-0.006
$SE_{it} * D_{it}$	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)
No. obs	873	873	855	855	855	855
Quadratic	No	Yes	No	Yes	No	Yes
Worst performing schools						
Treatment effect - Private	0.018	-0.068	0.003	-0.068	-0.010	-0.090
	(0.12)	(0.20)	(0.12)	(0.20)	(0.12)	(0.20)
Interaction term	-0.011	-0.011	0.000	0.000	0.046	0.048
$SE_{it} * D_{it}$	((0.01)	(0.01)	(0.00)	(0.00)	(0.03)	(0.03)
No. obs	373	373	372	372	373	373
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	-0.014	-0.041	-0.012	-0.039	-0.011	-0.038
	(0.03)	(0.06)	(0.03)	(0.06)	(0.03)	(0.06)
Interaction term	0.002	0.002	0.000	0.000	0.003	0.003
$SE_{it} * D_{it}$	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
No. obs	5774	5774	5743	5743	5743	5743
Quadratic	No	Yes	No	Yes	No	Yes

The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and pre enrollment data. All regressions use the baseline specification where the 50 mark cutoff is used. In column (1) the socio-economic variable controlled for is the Gini coefficient at the level of municipalities in 2000. In column (2), the variable used is the income per capita and in column (3), the average years of education of each municipality. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations.
Standard errors in parentheses
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Figure 1: Examples of school rankings published in the media: Newspaper Estado de São Paulo (Enem 2008), news website globo.com (Enem 2008) and Exame magazine (Enem 2013)

ESPECIAL
Tudo sobre o ENEM

Filtre por estado... administração... localização... ...ou busque por instituição

Todos | Todas | Todas | ok

(não use acentos)

Clique sobre as colunas para reordenar o quadro

instituição de ensino	UF	município	admin.	localização	nota*
COL DE SAO BENTO	RJ	Rio de Janeiro	Privada	Urbana	80,58
COLEGIO BERNOULLI	MG	Belo Horizonte	Privada	Urbana	77,38
COL DE APLICACAO DA UFV - COLUNI	MG	Viçosa	Federal	Urbana	76,66
COL STO ANTONIO	MG	Belo Horizonte	Privada	Urbana	76,43
COLEGIO HELYOS	BA	Feira de Santana	Privada	Urbana	76,34
COLEGIO WR	GO	Goiânia	Privada	Urbana	76,26
COLEGIO SANTO INACIO	RJ	Rio de Janeiro	Privada	Urbana	76,09
JUAREZ DE SIQUEIRA BRITTO WANDERLEY ENG CC	SP	São José Dos Campos	Privada	Urbana	76,02
VERTICE COLEGIO UNID II	SP	São Paulo	Privada	Urbana	75,97
COLEGIO SANTO AGOSTINHO	RJ	Rio de Janeiro	Privada	Urbana	75,97
COLEGIO SANTO INACIO	RJ	Rio de Janeiro	Privada	Urbana	75,92
BANDEIRANTES COLEGIO EFM	SP	São Paulo	Privada	Urbana	75,86
COLEGUIUM - ENSINO FUNDAMENTAL E MEDIO	MG	Belo Horizonte	Privada	Urbana	75,71
COLEGIO DE APLICACAO DO CE DA UFPE	PE	Recife	Federal	Urbana	75,68

Confira as melhores e as piores escolas do Brasil

Brasil | Rio de Janeiro | Brasília | São Paulo | São Paulo (Capital)

GERAL: MELHORES PIORES | REDE PARTICULAR: MELHORES PIORES | REDE PÚBLICA: MELHORES PIORES

BRASIL AS 20 MELHORES

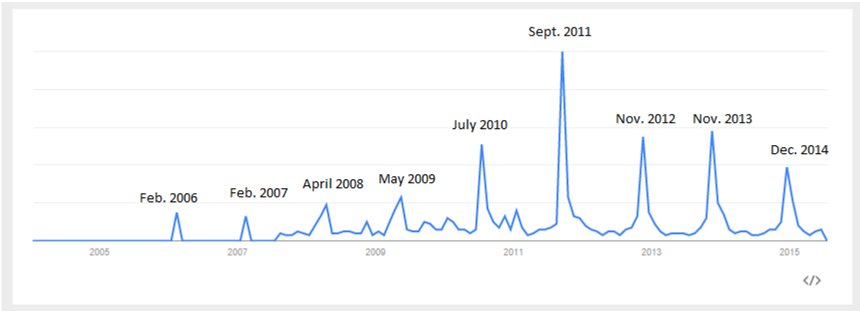
LEGENDA

Ranking elaborado com base na média total (prova objetiva e redação) com correção de participação (=simulação da nota como se 100% dos alunos tivessem participado)

1	RJ	Rio de Janeiro	Colégio de São Bento	Particular	Urbana	EMR	80,58
2	MG	Belo Horizonte	Colégio Bernoulli	Particular	Urbana	EMR	77,38
3	MG	Viçosa	Col. de Aplicação da UFV - COLUNI	Federal	Urbana	EMR	76,66
4	MG	Belo Horizonte	Col. Sto. Antônio	Particular	Urbana	EMR	76,43
5	BA	Feira de Santana	Colégio Helyos	Particular	Urbana	EMR	76,34
6	GO	Goiânia	Colégio WR	Particular	Urbana	EMR	76,26
7	RJ	Rio de Janeiro	Colégio Santo Inácio	Particular	Urbana	EMR	76,09
8	SP	São José dos Campos	Colégio Eng. Juarez de Siqueira Britto Wanderley	Particular	Urbana	EMR	76,02
9	RJ	Rio de Janeiro	Colégio Santo Agostinho	Particular	Urbana	EMR	75,97
10	SP	São Paulo	Colégio Vértice Unidade II	Particular	Urbana	EMR	75,97
11	RJ	Rio de Janeiro	Colégio Santo Inácio	Particular	Urbana	EMR e EJA	75,92
12	SP	São Paulo	Colégio Bandeirantes	Particular	Urbana	EMR	75,86
13	MG	Belo Horizonte	Coleguium - Ensino Fundamental e Médio	Particular	Urbana	EMR	75,71
14	PE	Recife	Colégio de Aplicação do CE da UFPE	Federal	Urbana	EMR	75,68
15	PI	Teresina	Inst. Dom Barreto	Particular	Urbana	EMR	75,5

ESTADO	CIDADE	PÚBLICA OU PRIVADA	NOME DA ESCOLA	MÉDIA DAS PROVAS OBJETIVAS	MÉDIA DA REDAÇÃO
SP	SAO PAULO	Privada	OBJETIVO COLEGIO INTEGRADO	741,94	804,55
MG	BELO HORIZONTE	Privada	COLEGIO BERNOULLI - UNIDADE LOURDES	722,64	792,53
RJ	RIO DE JANEIRO	Privada	COLEGIO E CURSO PONTO DE ENSINO	720,02	762,67
PI	TERESINA	Privada	INST DOM BARRETO	713,39	805,33
CE	FORTALEZA	Privada	ARI DE SA CAVALCANTE COLEGIO - MAJOR FACUNDO	710,67	808,70
DF	BRASILIA	Privada	COL OLIMPO	701,23	775,76
BA	FEIRA DE SANTANA	Privada	COLEGIO HELYOS	689,68	811,06

Figure 2: Evolution of Internet searches on Enem rankings over time using Goggle trends



The chart shows the relative importance of searches related to Enem rankings compared to total searches over the period, and therefore does not represent absolute values.

Figure 3: Frequency of schools by number of Enem takers

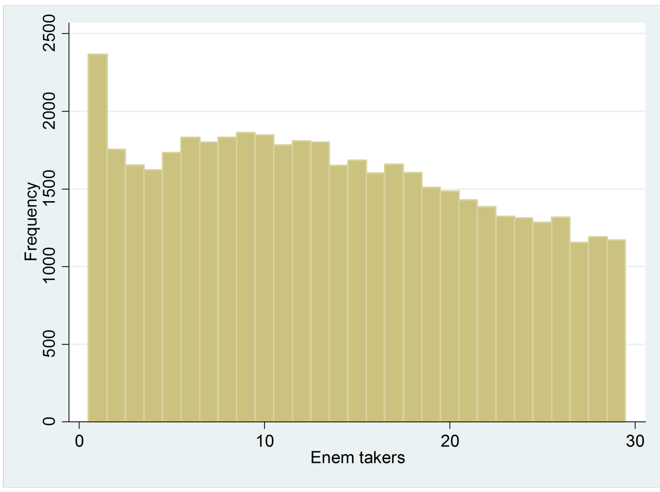
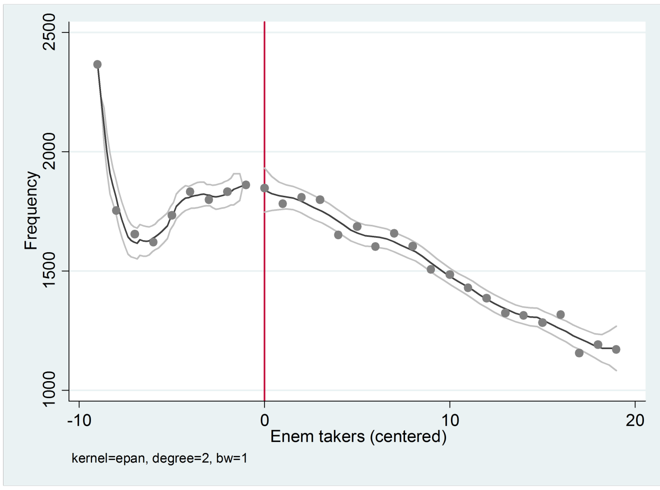


Figure 4: Local polynomial fit of frequency of schools by number of Enem takers



Grey lines represent the 95% confidence interval. The Epanechnikov kernel function was used, a polynomial degree of 2 and bandwidth of 1. The assignment variable has been centered so that the discontinuity is at 0.

Figure 5: Distribution of grades of public and private schools

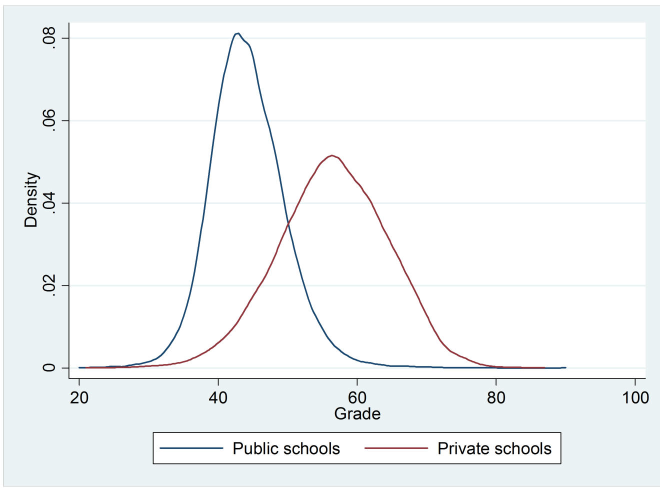
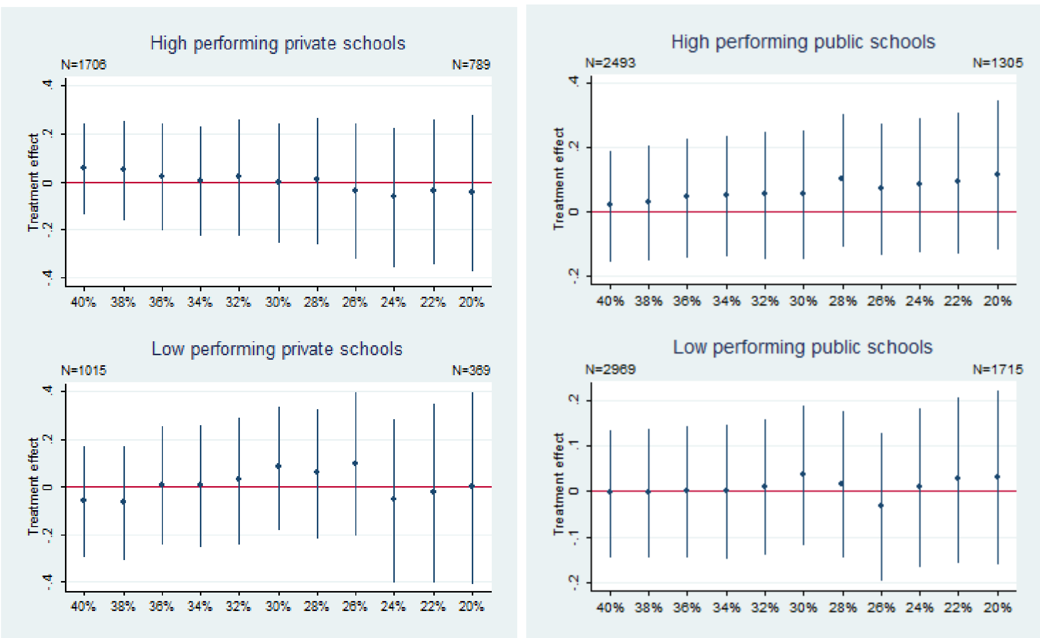


Figure 6: Dummy coefficient plot for cutoffs based on different percentiles



These graphs represent the dummy coefficient values obtained in regressions using different cutoffs for splitting the sample. For the best performing schools (above), the first point shows the coefficient obtained when we consider the 40% best schools, and the last point shows the coefficient obtained when we consider the 20% best schools. For the worse performing schools (below), the first point shows the coefficient obtained when we consider the 40% worse schools, and the last point shows the coefficient obtained when we consider the 20% worse schools. All estimations include a quadratic term for the number of test takers.

Appendix

A Additional Tables

Table A.1: Jumps in covariates estimation – private schools

	Avg. Grade		% low income*		% black		% father w/ high school	
Dummy coef.	-0.042	0.604	0.014	0.015	0.002	0.006	0.008	0.027
(Enem takers>=10)	(0.30)	(0.53)	(0.01)	(0.02)	(0.00)	(0.01)	(0.01)	(0.02)
Quadratic	No	Yes	No	Yes	No	Yes	No	Yes
Avg. at left of discontinuity	54		0.21		0.06		0.51	
No. obs.	8352		8352		8352		8352	

	% mother w/ high school		Log enrollment 1st year of high school		Log enrollment high school (total)	
Dummy coef.	0.005	0.021	-0.032	-0.119	-0.009	-0.085
(Enem takers>=10)	(0.01)	(0.02)	(0.04)	(0.07)	(0.03)	(0.07)
Quadratic	No	Yes	No	Yes	No	Yes
Avg. at left of discontinuity	0.60		3		4	
No. obs.	8352		4031		4117	

Controls are year dummies. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations.
*Includes the two lower categories of revenue among seven categories in the survey, equivalent to up to 2 minimum wages between 2005 and 2008 (600 to 830 BRL or 200 to 270 USD approximately)
Standard errors in parentheses
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table A.2: Jumps in covariates estimation – public schools

	Avg. Grade		% low income*		% black		% father w/ high school	
Dummy coef.	-0.242	-0.076	-0.007	-0.008	0.006	-0.007	0.000	-0.007
(Enem takers>=10)	(0.19)	(0.37)	(0.01)	(0.02)	(0.00)	(0.01)	(0.00)	(0.01)
Quadratic	No	Yes	No	Yes	No	Yes	No	Yes
Avg. at left of discontinuity	42.1		0.60		0.09		0.08	
No. obs.	11283		11284		11284		11284	

	% mother w/ high school		Log enrollment 1st year of high school		Log enrollment high school (total)	
Dummy coef.	0.006	-0.005	0.001	-0.121*	-0.010	-0.150**
(Enem takers>=10)	(0.01)	(0.01)	(0.03)	(0.07)	(0.03)	(0.07)
Quadratic	No	Yes	No	Yes	No	Yes
Avg. at left of discontinuity	0.14		4.1		4.9	
No. obs.	11283		7000		4117	

Controls are year dummies. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations.
*Includes the two lower categories of revenue among seven categories in the survey, equivalent to up to 2 minimum wages between 2005 and 2008 (600 to 830 BRL or 200 to 270 USD approximately)
Standard errors in parentheses
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Table A.3: Baseline estimates - window of 4-16 Enem takers

	50 mark		40% top/bottom schools		20% top/bottom schools	
Best performing schools	(1)		(2)		(3)	
Treatment effect - Private	0.018	-0.001	0.026	0.011	0.045	0.027
	(0.03)	(0.06)	(0.05)	(0.08)	(0.07)	(0.14)
No. obs	3703	3703	1975	1975	916	916
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	0.103	0.127	0.044	-0.005	0.093*	0.158
	(0.07)	(0.13)	(0.04)	(0.07)	(0.06)	(0.10)
No. obs	997	997	2866	2866	1497	1497
Quadratic	No	Yes	No	Yes	No	Yes
Worst performing schools						
Treatment effect - Private	-0.018	-0.036	-0.064	-0.046	-0.023	0.027
	(0.11)	(0.18)	(0.06)	(0.11)	(0.11)	(0.18)
No. obs	447	447	1201	1201	441	441
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	-0.015	-0.026	-0.005	-0.007	-0.01	0.008
	(0.03)	(0.05)	(0.04)	(0.06)	(0.05)	(0.09)
No. obs	6676	6676	3432	3432	1987	1987
Quadratic	No	Yes	No	Yes	No	Yes

The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in t=0. In column (1) the best and worst schools are separated using the 50 mark cutoff. In column (2) a ranking of schools is made inside each category (private or public) and the best and worst schools are the 40% higher and lower performing schools. In column (3) the best and worst schools are the 20% higher and lower performing schools inside each category. A symmetrical window of data around the discontinuity of 4-16 Enem takers is used in the estimations.
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Table A.4: Baseline estimates - window of 3-17 Enem takers

	50 mark		40% top/bottom schools		20% top/bottom schools	
Best performing schools	(1)		(2)		(3)	
Treatment effect - Private	0.023	-0.016	0.039	-0.007	0.057	0.016
	(0.03)	(0.05)	(0.04)	(0.07)	(0.07)	(0.12)
No. obs	4172	4172	2250	2250	1057	1057
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	0.103*	0.116	0.054	0.001	0.103**	0.104
	(0.06)	(0.11)	(0.04)	(0.06)	(0.05)	(0.09)
No. obs	1157	1157	3298	3298	1734	1734
Quadratic	No	Yes	No	Yes	No	Yes
Worst performing schools						
Treatment effect - Private	-0.044	-0.001	-0.082	-0.047	-0.047	-0.031
	(0.10)	(0.17)	(0.06)	(0.10)	(0.10)	(0.17)
No. obs	515	515	1359	1359	509	509
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	-0.013	-0.029	-0.005	-0.011	-0.011	0.003
	(0.02)	(0.04)	(0.03)	(0.05)	(0.04)	(0.08)
No. obs	7663	7663	3927	3927	2266	2266
Quadratic	No	Yes	No	Yes	No	Yes

The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in t=0. In column (1) the best and worst schools are separated using the 50 mark cutoff. In column (2) a ranking of schools is made inside each category (private or public) and the best and worst schools are the 40% higher and lower performing schools. In column (3) the best and worst schools are the 20% higher and lower performing schools inside each category. A symmetrical window of data around the discontinuity of 3-17 Enem takers is used in the estimations.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table A.5: Baseline estimates - total high school enrollment as outcome variable

	50 mark		40% top/bottom schools		20% top/bottom schools	
Best performing schools	(1)		(2)		(3)	
Treatment effect - Private	-0.013	0.05	0.022	0.106	0.065	0.14
	(0.03)	(0.06)	(0.04)	(0.09)	(0.08)	(0.16)
No. obs	3253	3253	1737	1737	801	801
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	0.086	0.046	0.004	-0.011	0.084	0.051
	(0.07)	(0.15)	(0.04)	(0.08)	(0.05)	(0.11)
No. obs	883	883	2515	2515	1318	1318
Quadratic	No	Yes	No	Yes	No	Yes
Worst performing schools						
Treatment effect - Private	-0.031	0.046	-0.078	0.032	-0.009	0.078
	(0.10)	(0.19)	(0.07)	(0.12)	(0.10)	(0.19)
No. obs	389	389	1049	1049	385	385
Quadratic	No	Yes	No	Yes	No	Yes
Treatment effect - Public	-0.015	-0.042	0.007	0.018	-0.006	-0.018
	(0.03)	(0.05)	(0.03)	(0.06)	(0.05)	(0.09)
No. obs	5837	5837	2999	2999	1730	1730
Quadratic	No	Yes	No	Yes	No	Yes

The dependent variable is the log of total enrollments in high school. Controls are year dummies and enrollment data in t=0. In column (1) the best and worst schools are separated using the 50 mark cutoff. In column (2) a ranking of schools is made inside each category (private or public) and the best and worst schools are the 40% higher and lower performing schools. In column (3) the best and worst schools are the 20% higher and lower performing schools inside each category. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations.
Standard errors in parentheses
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table A.6: Baseline estimates – State local rankings

	50% top/bottom schools	
Best performing schools	(1)	
Treatment effect - Private	0.001	0.018
	(0.04)	(0.06)
No. obs	3441	3441
Quadratic	No	Yes
Treatment effect - Public	0.028	-0.053
	(0.05)	(0.09)
No. obs	2376	2376
Quadratic	No	Yes
Worst performing schools		
Treatment effect - Private	-0.013	-0.068
	(0.26)	(0.42)
No. obs	126	126
Quadratic	No	Yes
Treatment effect - Public	-0.018	-0.003
	(0.03)	(0.06)
No. obs	4271	4271
Quadratic	No	Yes

The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in t=0. A local ranking of schools is made for each state, and the best and worst schools are the 50% higher and lower performing schools in each state. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations.
Standard errors in parentheses
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$