UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

ESSAYS IN THE ECONOMICS OF EDUCATION IN BRAZIL

ENSAIOS EM ECONOMIA DA EDUCAÇÃO NO BRASIL

Andrea Gruenwald Lépine

Orientador: Prof. Dr. Ricardo de Abreu Madeira

SÃO PAULO
2016

Prof. Dr. Marco Antonio Zago
Reitor da Universidade de São Paulo

Prof. Dr. Adalberto Américo Fischmann
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Hélio Nogueira da Cruz
Chefe do Departamento de Economia

Prof. Dr. Márcio Issao Nakane
Coordenador do Programa de Pós-Graduação em Economia

ANDREA GRUENWALD LÉPINE

ESSAYS IN THE ECONOMICS OF EDUCATION IN BRAZIL

ENSAIOS EM ECONOMIA DA EDUCAÇÃO NO BRASIL

Tese apresentada ao Programa de Pós-Graduação do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito parcial para obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Ricardo de Abreu Madeira

Versão Corrigida

(versão original disponível na Biblioteca da Faculdade de Economia, Administração e Contabilidade)

SÃO PAULO

2016

FICHA CATALOGRÁFICA
Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Lépine, Andrea Gruenwald.
Essays in the economics of education in Brazil / Andrea Gruenwald Lépine. -- São Paulo, 2016.
117 p.

Tese (Doutorado) – Universidade de São Paulo, 2016.
Orientador: Ricardo de Abreu Madeira.

1. Economia da educação. 2. Responsabilização das escolas. 3. Incentivos. 4. Ensino superior. I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 370.193

À minha família.

AGRADECIMENTOS

Em primeiro lugar agradeço ao meu orientador, Professor Ricardo Madeira, pelo apoio durante estes anos de tese, pela paciência, confiança, e grande liberdade que me deu na realização deste trabalho de pesquisa. Seus conselhos foram muito valiosos.

Ao Professor Marcos Rangel, que me guiou nos primeiros passos para entrar no programa de doutorado na FEA, e me orientou no começo da tese. Sua ajuda e confiança foram fundamentais, assim como seus conselhos na definição do projeto de tese e durante a pesquisa.

À Professora Fernanda Estevan, pelos importantes comentários feitos durante o exame de qualificação, assim como pela disponibilidade em conversar. Agradeço também aos outros professores da pós-graduação da FEA pelos ensinamentos, e em particular ao Professor Fernando Botelho pela ajuda com bases de dados.

Ao Departamento de Economia e ao pessoal da Secretaria de Pós-Graduação da FEA por todo o apoio prestado.

Ao CNPq e Fapesp pelo apoio financeiro que permitiu a realização desta tese, e permitiu que eu participasse em conferências que foram fundamentais para o amadurecimento deste trabalho.

Gostaria também de agradecer aos colegas da pós, que graças ao seu bom humor e solidariedade ajudaram a tornar mais leves os desafios desta etapa: Ana, Dario, Juliana, Rodrigo, Natália e vários outros.

Enfim, este trabalho só foi possível graças ao apoio de minha família durante esta desafiante etapa da minha vida. Agradeço aos meus pais, Ruth e Jacques, por sempre me incentivar e estar presentes nos momentos mais difíceis desta jornada. Sem a imensa ajuda da minha mãe, sem dúvida teria sido impossível conciliar a tese e a maternidade. Ao meu marido, Sebastián, pela compreensão, carinho e otimismo. À minha avó, Nair, por sempre acreditar em mim e me encorajar.

RESUMO

Esta tese é composta por três ensaios independentes em economia da educação, que têm como objetivo avaliar o impacto de diferentes políticas públicas no Brasil.

O primeiro capítulo estuda os efeitos do *Prouni*, um programa que oferece bolsas para estudantes de baixa renda no ensino superior. A metodologia usada para lidar com efeitos de seleção é a de *propensity score matching*, com base em características observáveis dos alunos e uma *proxy* para desempenho inicial. Os resultados, que são robustos a diferentes especificações, mostram que os bolsistas têm melhor desempenho e levam menos tempo para chegar ao último ano de faculdade do que alunos comparáveis. Os efeitos estimados são maiores para bolsistas integrais do que para bolsistas parciais, sugerindo que o valor da bolsa é relevante. Os resultados também indicam que os bolsistas integrais têm probabilidade menor de trabalhar durante a faculdade.

O segundo capítulo apresenta evidências empíricas relativas ao efeito do programa de incentivos para professores do Estado de São Paulo, no qual foram distribuídos bônus de grupo para professores e funcionários em função da melhora do desempenho dos alunos. Neste capítulo é usado o método de diferenças em diferenças e de diferenças triplas. Os resultados sugerem que o programa teve efeitos positivos sobre o desempenho dos alunos, no entanto, os efeitos encontrados variam entre séries e matérias. Embora poderia esperar-se que em escolas com um maior número de professores efeitos de *free-riding* sejam mais fortes e limitem o efeito da política, este efeito não parece ser importante. O estudo mostra também que escolas com desempenho inicialmente baixo apresentaram melhoras mais importantes do que a média.

O terceiro capítulo analisa se a divulgação de informações sobre a qualidade das escolas afeta as escolhas dos alunos. Mais especificamente, procura-se saber se a divulgação das notas obtidas no *Enem* (Exame Nacional do Ensino Médio) afetou o volume de matrículas em escolas com alto e baixo desempenho. A metodologia usada é a de regressão descontínua, com base em uma regra exógena que determinou que somente as escolas com um número mínimo de alunos que fizeram a prova teriam seus resultados divulgados. Os resultados indicam que a divulgação de notas a nível de escolas não teve impacto significativo sobre as decisões de matrícula, tanto no caso de escolas privadas como públicas. Os resultados se mantêm quando é levado em consideração o grau de concorrência enfrentado pelas escolas, ou o seu ambiente socioeconômico.

ABSTRACT

This thesis is composed by three independent essays in economics of education, which aim to assess the impact of different public policies in Brazil.

The first chapter studies the effects of a government scholarship program for low-income college students, the *Prouni*. Propensity score matching based on observable student characteristics and a proxy for previous student performance is used to deal with selection effects. The results are robust across different specifications, and suggest that students who received a scholarship perform better and take less time to reach their final year of college than comparable students. The estimated effects are higher among students with full scholarships than for students with partial scholarships, indicating that the amount of aid received matters. Results also indicate that full scholarship recipients also have a lower probability of working while in college.

The second chapter provides evidence on a large-scale pay for performance program in the state of São Paulo, which awarded group bonuses to school teachers and staff conditional on improvements in student performance. Results using a difference-in-differences and triple-differences framework show that the program had overall positive effects on student performance, although improvements vary across grades and subjects. Although it could be expected that free-riding effects increase with the number of teachers in schools, thereby limiting the impact of the program, this effect seems to be modest. Results also show that initially low-performing schools improved much more than the average.

The third chapter analyzes whether the provision of information on school quality affects students' school choices. More specifically, it explores 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. A sharp regression discontinuity design is used, taking advantage of the fact that an exogenous rule determined that only schools with a minimum number of test-takers would have their results published. The results show that the disclosure of school grades did not significantly affect enrollment decisions, in neither private nor public schools. The findings remain unchanged when controlling for the degree of competition faced by schools or their socio-economic environment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ATT - Average treatment effect on the treated

DD - Difference-in-differences

DDD - Triple-differences

Enade - Exame Nacional de Desempenho de Estudantes

Enem - Exame Nacional do Ensino Médio

Fies - Programa de Financiamento Estudantil

Ideb - Índice de Desenvolvimento da Educação Básica

Idesp - Índice de Desenvolvimento da Educação do Estado de São Paulo

Inep - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

IPEA - Instituto de Pesquisa Econômica Aplicada

OECD - Organisation for Economic Cooperation and Development

OLS - Ordinary Least Squares

PISA - Program for International Student Assessment

Prouni - Programa Universidade Para Todos

PSM - Propensity Score Matching

Saresp - Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo

SEMESP - Sindicato das Mantenedoras de Ensino Superior

Sisu - Sistema de Seleção Unificada

UNDP - United Nations Development Programme

INTRODUCTION

Brazil has made great improvements in access to education over the last fifteen years, and school enrollment is now close to that of high-income countries. However, education quality remains low even compared to countries with similar levels of per capita income, as evidenced by international student assessments such as PISA (*Program for International Student Assessment*). In particular, 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. This, combined with the fact that free public universities in Brazil are generally of higher quality and more difficult to access, means it is particularly difficult for disadvantaged students to access good quality higher education, which perpetuates inequalities and limits social mobility.

Policies aiming at improving the quality of education, increasing school accountability, and more generally increasing opportunities for low-income students have been at the center of debates recently. Different educational policies have been implemented over the past years at the municipal, state or national level. In this context, rigorous evidence on the impact of these policies is of fundamental importance, so that they can be better designed and better targeted in the future.

This thesis is composed by three independent empirical essays in the economics of education in Brazil, which study the impact of different policies affecting educational outcomes.

The first evaluates the effects of scholarships for disadvantaged college students in Brazil through the *Prouni* program (*Programa Universidade para Todos*). As higher education becomes increasingly important in the labor market, student financial aid has a key role in increasing access and improving college outcomes for disadvantaged students. However, empirical evidence on this type of policy in the context of developing countries is extremely limited. Propensity score matching is used to create a counterfactual of students who did not receive scholarships, and results suggest scholarship recipients perform better than similar students and take less time to reach their final year of college. Although the identifying hypothesis is based on the relatively strong assumption of selection on observables, the findings are robust to alternative specifications and pass different robustness tests. The estimated effects are also stronger for students receiving a full scholarship covering 100% of tuition than for students receiving a partial scholarship covering 50% of tuition, suggesting that the amount of aid matters.

The second chapter evaluates the impact of a teacher incentive program in the State of São Paulo, which awarded group bonuses conditional on improvements in student performance. Despite being highly controversial, monetary incentives for teachers are an increasingly popular policy, although the effects of this type of policy are not clear from a theoretical perspective and empirical evidence is limited so far. To assess the impact of the policy, a difference-in-differences and a triple-differences framework are used. The results suggest the program had overall positive effects in performance, although improvements vary by grade and subject. Overall, the analysis does not suggest the presence of strong free-riding effects. Heterogeneous effects according to initial school performance are also estimated and point to initially lower-performing schools improving much more than the average.

The third chapter looks at the effects of providing public information on school quality, through the publication of school-level test scores obtained at the *Enem (Exame Nacional do Ensino Médio)*, a standardized test taken by high school students in Brazil. Although it could be expected that this type of information would affect students' school choices, the literature suggests this is not necessarily the case, as preferences regarding school quality are heterogeneous. To establish causality, a sharp regression discontinuity design is used by taking advantage of an exogenous rule determining that only schools with a minimum number of test-takers would have their results published. The findings show that the disclosure of school grades did not result in significant changes in enrollment choices, in neither private nor public schools. The findings remain unchanged when controlling for the degree of competition faced by schools or their socio-economic environment.

1 Effects of Scholarships for Disadvantaged College Students in Brazil

1.1 Introduction

Access to tertiary education in developing countries is relatively low compared to high-income countries. Among Latin American countries for which OECD data is available[1], the most recent estimates show that on average 19% of the population aged 25-64 had completed higher education, compared to nearly 35% for OECD member countries. In Brazil, this rate was only 14% despite the fact that returns to higher education are particularly strong[2]. One of the reasons behind this phenomenon is the cost of higher education, combined with the presence of credit constraints. A World Bank study by Murakami and Blom (2008) shows for example that the costs associated with higher education (which include the cost of living) in terms of per capita income are much higher in Latin America than in high-income countries.

In addition to obstacles related to initial enrollment, many students drop out of college before completing their studies because of poor performance or financial difficulties, and many take longer than expected to finish. Available statistics show almost a third of students who enroll in tertiary education do not graduate in OECD countries[3], and estimates provided by Binelli, Meghir and Menezes-Filho (2008) and other sources point to similar rates for private college students in Brazil.

Student financial aid in the form of grants or scholarships, which still plays a limited role in developing countries, can potentially help disadvantaged students access and successfully complete higher education. In addition to reducing the cost of higher education, student aid might allow students to devote more time to their studies and increase their performance by reducing the need to work while in college.

The available empirical evidence on the effectiveness of student aid, mainly focused on programs in the United States, suggests that it positively affects student enrollment and persistence, although there is a large variation in the size of the estimated impacts. In a well-known paper, Dynarski (2003) uses difference-in-differences to estimate the impact

[1] Organisation for Economic Cooperation and Development. Latin American countries with available data include Chile, Brazil, Mexico, Argentina, Colombia and Costa Rica (data from 2013 and 2014).

[2] Those who completed higher education in Brazil were estimated to earn on average 2.5 times more than those with only upper secondary education, compared to a rate of 1.6 for OECD countries.

[3] Education at a Glance 2013: OECD Indicators

of the elimination of a student aid program in the United States, The Social Security Student Benefit Program, and finds that the program had positive effects on attendance and completion. A similar approach is used by Cornwell, Mustard and Sridhar (2006) who analyze the effects of the HOPE Program in Georgia, which awarded merit-based scholarships for students attending in-state higher education institutions as well as grants for students attending technical schools. They find that the program increased enrollment, especially for black students. In another study, Kane (2003) explores discontinuities on the eligibility criteria of a student aid program in California (the Cal Grant Program) and also finds positive effects on enrollment among students who had already applied for financial aid. A more recent paper by Angrist, Autor, Hudson and Pallais (2014) uses a randomized evaluation to study the effects of grants based on grade and financial need distributed to high school seniors in Nebraska, and finds increases in enrollment as well as positive effects on persistence, with larger gains among the most disadvantaged students. Outside of the United States, Dearden, Fitzsimons and Wyness (2014) and Nielsen, Sørensen and Taber (2008) find evidence that student aid positively affects enrollment using data from the UK and Denmark, respectively.

To my knowledge, the only paper analyzing the effects of student aid in the context of a developing country is a study by Canton and Blom (2004), who try to estimate the impacts of a student loan program in Mexico (SOFES). But although they find positive effects on enrollment and performance, the study faces some important methodological limitations and it is not clear whether its findings would generalize to the case of non-refundable aid such as grants and scholarships.

This paper contributes to the literature on higher education and student financial aid by assessing the effects of a scholarship program for disadvantaged students in Brazil, the *Prouni* (*Programa Universidade Para Todos*), on the outcomes of students in their final year of college. In particular, I look at how the scholarship affects student performance and the duration of studies. Despite the importance of the acquisition of skills during higher education, there is very limited evidence on how scholarships affect actual student learning and performance. In order to deal with selection effects, I use propensity score matching to create a counterfactual group of students with similar socio-economic characteristics who did not receive scholarships. I take advantage of the fact that I can observe both first-year and final-year students' performance on a specific knowledge test related to their field of study, as well as on a general knowledge test. The data show that while specific knowledge increases throughout college, general knowledge is much more stable across time, which allows me to use the general knowledge grade as a proxy for previous student performance before enrolling in college.

Results show that final-year students who have received a *Prouni* scholarship perform better than similar students who have not received a scholarship, and take less time to reach the final year of college. These effects are stronger for students who received full scholarships than for students who received partial scholarships (covering 100% and 50% of tuition respectively), suggesting that the amount of aid received also matters. Results also show that scholarship recipients report studying more on average, and that those who received a full scholarship have a lower probability of working while in college, though this is not observed for partial scholarship recipients. The findings are robust to alternative specifications and pass different robustness tests.

The remainder of this paper is organized as follows. Section 1.2 provides background information on the higher education system in Brazil and the *Prouni* program. Section 1.3 presents the data and details the methodology. Section 1.4 shows the main results, and a few robustness tests are performed in Section 1.5. Section 1.6 concludes.

1.2 Background

1.2.1 Higher Education in Brazil

Higher education in Brazil is provided by both fee-paying private institutions, which account for the large majority of enrollments, and free public institutions. According to data from the Ministry of Education, around 25% of enrollments in 2014 were in public institutions. Although college attainment is still relatively low compared to high-income countries (only 14% of people aged 25-64 had completed higher education in 2014 according to data from the OECD, compared to an average of 35% for OECD member countries), the number of enrollments in tertiary education has been increasing significantly in recent years, and rose by around 75% between 2005 and 2014[4].

A specific characteristic of the Brazilian higher education system is that public institutions are generally of higher quality and more selective than private ones. According to Binelli et al. (2008), there were on average 9 applicants for each place at a public institution in 2003, while this ratio was 1.5 in private institutions. The opposite is true for basic education, where private schools generally outperform public schools at standardized tests. It is therefore particularly difficult for students from disadvantaged backgrounds who attended public schools to access the best public universities, and inequalities persist throughout higher education. Student dropout is another important issue, especially in private institutions. Binelli et al. (2008) estimate these rates at around 20% for public institutions and 33% for private ones in 2002, and a more recent study by SEMESP[5] in 2015 estimated the dropout rate at around 27% for private higher education institutions

[4]Data from the Ministry of Education

[5]*Sindicato das Mantenedoras de Ensino Superior*, an organization that provides services for private higher education institutions in Brazil.

and 18% for public institutions.

The admission process in Brazilian higher education institutions is mainly decentralized, and most colleges have their own admission criteria and entrance tests. In the case of public institutions, however, there is a tendency towards the centralization of admissions with the creation in 2010 of a single platform to select and allocate students to participating institutions (*Sisu*[6]). Overall, the difficulty of admissions can vary greatly, and the selection of students can be based on performance at specific entrance tests, on test scores obtained at the *Enem* (*Exame Nacional do Ensino Médio* - a standardized test aimed at students finishing high school), and on interviews and curriculum examinations, among others. Students are required to choose their field of specialization before applying and enrolling, and once students are admitted, change is usually difficult and requires going through the selection process again.

Several initiatives have been taken by the federal government in recent years to increase and democratize access to higher education, such as the introduction of scholarships for disadvantaged students (*Prouni*), the creation of a loan program for disadvantaged students (*Fies - Programa de Financiamento Estudantil*), and quotas for students from public schools. Below, I describe the *Prouni* program in further detail.

1.2.2 The ***Prouni*** Program

The *Prouni* (*Programa Universidade Para Todos*) is a program created in 2005 by the Brazilian Federal Government that offers scholarships to students enrolled in private higher education institutions. Two types of scholarships are available: full scholarships that cover 100% of tuition costs, and partial scholarships that cover 50% of tuition. Private higher education institutions participating in the program agree to reserve a certain number of spots for *Prouni* students, and in return benefit from tax exemptions[7]. The evolution of the number of *Prouni* scholarships between 2005 and 2014 is shown in Figure 1.1. In 2014, over 300,000 scholarships were awarded, of which nearly 70% were full scholarships.

In order to be eligible, students must satisfy the following criteria: (1) they must have previously attended either a public high school, or a private high school receiving a full scholarship; (2) their household income must not exceed 1.5 minimum wages per capita if applying for a full scholarship or 3 minimum wages per capita if applying for a partial scholarship[8]. Teachers from public schools studying Pedagogy or studying for a teaching

[6] *Sistema de Seleção Unificada*

[7] The number of *Prouni* spots available in each degree program is a function of the total number of spots available, although the rules allow for some flexibility.

[8] In 2016 the minimum wage in Brazil was 880 Reais, equivalent to 250 USD approximately.

degree, and students with disabilities are also eligible, regardless of their income or school attended.

Students who wish to apply for a *Prouni* scholarship need to go through an online centralized selection process that happens twice a year, where they are allowed to choose up to two different degree programs[9]. Candidates are ranked according to their *Enem* score, and pre-selected based on the program's minimum grade requirements and the number of spots available (information on the availability of spots is updated in real time). Higher education institutions can additionally require pre-selected students to go through their own selection process. The program's rules have changed slightly over the years and in 2009 for example, the requirement that students score above a minimum threshold on the *Enem* exam was added - although the fixed threshold is relatively low and more than half the students taking the *Enem* score above this minimum grade. Once a student is awarded a *Prouni* scholarship, its validity is reviewed on a term basis and students need to pass at least 75% of the classes taken in a given term in order to keep the scholarship.

1.3 Data and Empirical Strategy

1.3.1 Data

This paper uses data from *Enade* (*Exame Nacional de Desempenho de Estudantes*), an exam taken each year by a sample of undergraduate students in higher education institutions in Brazil. The *Enade*'s purpose is to assess the quality of undergraduate degree programs in the country, and taking the exam is mandatory although it has low direct stakes for students. The *Enade* assesses both students' performance in their specific field of study and in general knowledge. In addition to student performance, the *Enade* database also provides detailed information on students' socio-economic background and on whether students have received scholarships or loans.

Each year, degree programs from different fields of study are assessed, and roughly the same group of fields is tested every three years, although new fields are added with time. Table 1.1 shows the fields of study tested in *Enade* between 2004 and 2010. As an example fields tested in 2004, the first year for which *Enade* data is available, were tested again in 2007 and 2010. Initially, the sample included both first-year and final-year students, but from 2011 onwards only final-year students were tested.

As students from different fields of study are likely to be significantly different in terms of both observable and unobservable characteristics, and as the difficulty of the exam may

[9]Students could choose up to five degree programs until 2009, but the number of choices was subsequently reduced.

vary from year to year, it makes sense to perform the analysis of the effects of receiving a *Prouni* scholarship separately by field and year. Fields included in the analysis need to have a sufficient number of students receiving both full and partial scholarships, as the effects of the two types of scholarships are assessed separately. However, the distribution of scholarship recipients between fields of study and years is very uneven. Traditional fields such as Law or Medicine typically attract a much higher number of students than others fields. Moreover, the number of *Prouni* students in the early years after the program was implemented was very limited in most fields.

In order to obtain a sufficient sample size for the analysis, I take, for each field and year, the minimum number of *Prouni* students of the two categories (full and partial)[10] and rank all combinations of field and year accordingly. I then restrict the analysis to the three fields that provide the largest number of observations. I exclude degrees in Pedagogy, which have specific scholarship attribution rules as mentioned earlier, and restrict the analysis to the period before 2011 as data for first-year students, which are used in the analysis, are not available after this period. According to these criteria, the three fields with the largest sample size are Management, Law and Accounting, obtained from the 2009 *Enade* database, all of which have at least 500 *Prouni* students of each category[11].

Participation in the *Prouni* program is a choice made by institutions, and those choosing to participate may have different characteristics than non-participating institutions. I therefore restrict the counterfactual of students not receiving scholarships to students from participating institutions, by excluding from the analysis degree programs in Management, Law and Accounting from colleges not participating in the *Prouni*[12] (programs from public institutions, which do not charge fees and are not eligible to participate in the program, are also excluded).

1.3.2 Descriptive Statistics

Table 1.2 presents descriptive statistics for final-year students in the three fields of study included in the analysis. Only students with available data on test scores are included, and aged between 17 and 50, as the number of *Prouni* students outside this age range is very limited. I also exclude students with disabilities from the analysis as they have different requirements for receiving *Prouni* scholarships. Overall, scholarship recipients represent about 15% of the total. The majority is composed by students who received a full *Prouni* scholarship (8-9%), followed by partial *Prouni* scholarship recipients (3-5%).

[10] This is done by pooling together all colleges offering degree programs in a given field.

[11] The analysis has also been done using the two next fields with the largest number of *Prouni* students, which provides similar results (not shown here).

[12] As I do not directly observe which private institutions officially participate in the *Prouni*, I only include in the analysis degree programs that have at least one *Prouni* student of either type. In 2009, only 19% of all degree programs in Management, Law and Accounting did not have any *Prouni* students

The share of students who received other types of scholarship is very limited[13].

Management is by far the most popular field of the three and accounts for the largest number of students, followed by Law and Accounting. The three fields of study differ somewhat regarding their student population. Law students come from a slightly more advantaged background, which translates into a lower proportion of those who attended public high schools, and a higher proportion of those whose parents finished high school. Law degree programs last longer than other degrees: while most non-technical degrees in Brazil have an average duration of 4 years, the Law degree normally takes 5 years. The large majority of final-year students works and studies in the evening period, although this share is lower among Law students.

An important point should be noted here. As scholarships are reviewed on a term basis and can be suspended if students' situation changes (for example if their family income increases above the maximum threshold or if they do not pass 75% of the classes in a given term), it is possible that some of the students who initially received a scholarship do not have it any longer at the time of the survey. The survey, however, only provides information on whether students have received a *Prouni* scholarship at some point during their studies. In order to get an idea of what this represents, I look at higher education census data for 2009 and 2010. In the three fields of study considered, around 11% of first-year students enrolled in 2009 who had a full *Prouni* scholarship lost it the following year, and for students who had a partial *Prouni* scholarship this share was around 14%.

In Table 1.3, similar statistics are shown by *Prouni* scholarship status and type of student (first-year or final-year). As expected, the share of low-income students and the share of students who attended a public high school increase with *Prouni* scholarship status[14]. Interestingly, *Prouni* students score higher in both specific and general knowledge tests, and final-year students who have received a *Prouni* scholarship take less time on average to reach their final year of studies. As these students have specific characteristics however, it not possible to establish any causal relations at this point.

When comparing first-year and final-year students, some interesting patterns emerge. Final-year students are more likely to be working and contributing to the household income, which is likely to explain the fact that the percentage of final-year students in the lowest income category is smaller. Comparing characteristics of first-year and final-year students that are fixed in time can also give some indications regarding student dropout. Although the number of final-year students is generally lower than the number of first-year students, there are no noticeable differences in most cases regarding mother and father

[13]Students who report receiving more than one type of scholarship represent less than 1% of the sample and are not included in the analysis.

[14]As mentioned earlier, given that the only available information is whether students have received a *Prouni* scholarship at some point during their studies, it is possible that some *Prouni* recipients do not satisfy the income eligibility requirements at the time the survey was applied.

education, gender, and race; although in some cases a decrease in the share of students who attended a public high school can be observed.

1.3.3 Empirical Strategy

Comparing the outcomes of *Prouni* scholarship recipients with those of other students is likely to be misleading, as students who receive scholarships have distinct characteristics, and in particular come from a more disadvantaged background. In order to create a counterfactual of students with similar observable characteristics, I use the propensity score matching method suggested by Rosenbaum and Rubin (1983). Following this approach, treated and non-treated individuals are matched based on each individual's probability of being treated (or propensity score), which is estimated using their observed covariates. If the hypothesis that assignment to treatment is random conditional on observable covariates is met, and if there is sufficient overlap in the distribution of propensity scores of treated and non-treated individuals, this method allows the estimation of the average treatment effect on the treated (ATT). Compared to OLS, propensity score matching has the advantage of restricting the comparison of the outcomes of treated and control individuals to those with similar characteristics, and of allowing for a more flexible functional form.

A limitation of this method is that even after matching individuals based on observed covariates, it is possible that treated and non-treated students still differ on unobserved characteristics, such as intrinsic motivation. To the extent that these unobserved characteristics affect both selection into the treatment and outcomes, estimates could be biased and therefore results should be interpreted with caution. However, the fact that eligibility for *Prouni* scholarships is based on discrete income thresholds and the fact that the number of scholarships available is limited means that, in practice, students who just miss the income or grade requirements but are otherwise very similar to *Prouni* recipients will be part of the control group.

Variables used to predict the probability of treatment should not only reflect students' socio-economic background but also prior performance, given that students compete for a limited number of *Prouni* scholarships and are selected based on *Enem* test scores. Failing to take that into account is likely to introduce bias in the estimations, given that previous ability is usually a strong determinant of current performance. Although I do not observe *Enem* test scores for students in the sample, I argue that students' performance in the general knowledge test can be used as a proxy for previous ability. Indeed, when comparing test scores of first-year and final-year students, the gap in specific knowledge is much larger than the gap in general knowledge, which seems to be less affected by college education (although it is likely to increase with time). These differences are shown in

Table 1.4, which presents OLS estimates of the effect of being a final-year student on test scores (first-year and final-year students are pooled together, and the coefficient of interest is obtained by including in the regressions a dummy=1 for final-year students). Estimates show that while final-year students score 0.6-0.7 standard deviations higher than first-year students on the specific knowledge test, this difference is much smaller when considering general knowledge grades, which increase by 0.1-0.2 standard deviations[15].

In order to implement the propensity score matching procedure, I first estimate the probability of being treated for full and partial *Prouni* scholarships separately, through a probit regression. I then use students' estimated propensity score to match treated and non-treated students, using different matching algorithms.

Before computing the results, I verify whether the matching was successful in creating similar treated and control groups in terms of observable characteristics. First, I check that covariates used in propensity score estimations are balanced among the two groups in the matched sample. When imbalances remain, I re-estimate the propensity score including interactions or quadratics in order to achieve balance, from the same group of covariates. I also visually inspect the distribution of propensity scores among the treated and control groups before and after matching, to verify that they become more similar. Finally, I re-estimate the propensity scores on the matched sample, and compare the corresponding pseudo R-squared from that obtained from the unmatched sample, as suggested by Sianesi (2004) and Caliendo and Kopeinig (2008). The idea behind this procedure is that when using the matched sample, the pseudo R-squared should be much lower as in this case treated and control individuals should have similar characteristics and covariates should not be able to explain participation as well as before. Once these steps are verified, I calculate the differences in outcomes of interest separately for students receiving a full *Prouni* scholarship and for students receiving a partial *Prouni* scholarship.

1.4 Results

Table 1.5 shows probit estimates of the probability of receiving a *Prouni* scholarship[16]. As expected, results show that income is negatively correlated with the probability of receiving a *Prouni* scholarship, while having attended a public high school increases the probability of being treated. Similarly, an individual's probability of receiving a scholarship increases with the grade obtained at the general knowledge test.

[15]The increase in general grades is more marked for Management students, however, as many colleges offer basic knowledge courses for first-year students.

[16]The eligibility criteria for receiving a Prouni scholarship are not perfectly observed in the Enade database. Only information on income ranges is available, and although information on the type of high school attended is provided, it does not say whether students attending private high schools received a scholarship.

Different tests verifying that the matching was successful are shown in the Appendix. Tables 1.A1 - 1.A6 show differences in covariates between treated and control individuals before and after matching. Only differences using the sample obtained through nearest neighbor matching are presented, as results using other matching algorithms are very similar[17]. While differences in the characteristics of treated and control individuals are statistically significant at the 5% level before matching in the majority of cases, no significant differences at the 5% level remain when considering the matched sample. Likewise, figures 1.A1-1.A3 show that the distribution of propensity scores, which was quite different between the treated and control groups before matching, becomes much more similar between groups in the matched sample. As a final check, Table 1.A7 shows that the pseudo R-squared is greatly reduced when re-estimating propensity scores using the matched sample, indicating that covariates have a lower explanatory power in this case.

Table 1.6 and 1.7 present estimates of the effect of receiving a *Prouni* scholarship on performance at the specific knowledge test (using standardized test scores), and on the number of years taken by students to reach their final year of college. Column 1 shows simple differences in outcomes between the treated and control groups, while Column 2 shows OLS estimates where the same variables used to estimate propensity scores are used as controls. In columns 3 to 6, estimates obtained through four different matching algorithms are presented. In column 3, nearest-neighbor matching assigns each treated student to the closest non-treated student in terms of the probability of being treated. In column 4 a similar procedure is used, but in this case the 5 closest individuals are used as matches. In column 5 radius matching is used, where all individuals whose propensity scores fall within a given distance of the propensity score of a given treated individual are used as matches. In kernel matching (column 6), all individuals of the control group are used but are given different weights so that the highest weights are given to individuals with the closest probability of treatment to a given treated individual. I make sure that the sort order of observations is random, as the order of observations could affect results when there are observations with identical propensity scores. Only observations in the common support are used[18].

Propensity score matching results indicate that final-year students who received a *Prouni* scholarship at some point during their studies perform significantly better than comparable students, and that this difference is higher for students receiving full *Prouni* scholarships than for students receiving partial scholarships. The largest effects are seen for Management and Accounting students. In these two fields, full *Prouni* recipients score 0.6-0.7 standard deviations higher than other students, while partial *Prouni* recipients

[17] Overall, covariates remain unbalanced after matching in 1% of cases, with significant differences at the 5% level.

[18] Observations that have a propensity score higher/lower than the maximum/minimum of the other group are excluded.

score 0.2-0.3 standard deviations higher. Law students who received a full *Prouni* scholarship score 0.4 standard deviations higher than other students, while those who received a *Prouni* scholarship score 0.1 standard deviations higher. In all three fields, scholarship recipients also take less time to reach their final year of college, with estimated effects ranging between 0.1 and 0.3 years less.

A possible interpretation of these results is that students who receive a scholarship become more motivated. It is also likely that scholarship recipients study more because they face the threat of loosing their scholarship if they do not pass at least 75% of classes taken in a given term. In order to explore the channels through which receiving a scholarship improves student outcomes, I also estimate the effects of receiving a scholarship on students' decision to work and on time spent studying. Table 1.8 presents the effects of receiving a scholarship on the decision of students to work. For simplicity, only propensity score matching results are shown. On average, the percentage of full *Prouni* recipients who work is between 3% and 7% lower than in the control group. However, partial *Prouni* recipients are not less likely to work than non-scholarship recipients. In Table 1.9, a similar exercise is performed where the outcome variable is a dummy which equals one if the student reports studying more than three hours a week. Both full and partial *Prouni* scholarship recipients report studying more than students from the control group, although this increase is higher in the former group.

1.5 Robustness Checks

In this section a few robustness checks are considered in order to assess the validity of the results.

In the first robustness test, I check whether differences in college quality are driving the results. This might be the case for example if the best or more selective colleges offer a disproportionately high number of spots for *Prouni* students. In order to deal with this possibility, I include measures of college quality in previous OLS and propensity score matching estimations presented in Tables 1.6 and 1.7. The results are shown in Tables 1.10 and 1.11. Column 1 presents OLS estimates where I control for the average test scores of first-year and final-year students' obtained at the general and specific knowledge tests. As an alternative, column 2 reports estimates obtained from OLS regressions including college dummies. Columns 3-6 present propensity score matching estimates where the same measures of college quality used in column 1 are included in the set of matching variables. Overall results are very similar as those obtained in the main specifications, although in some cases the estimated effect of receiving a scholarship on performance is slightly reduced compared to previous estimates.

In the second robustness test, I re-estimate propensity scores using a logit model instead of a probit model, before performing the same matching procedures as previously. The results obtained are very similar, and are not shown here.

Selection on observables does not account for characteristics such as intrinsic motivation, which might be a determinant of scholarship attribution and also be correlated with test scores. Intrinsic motivation might help explain, for example, why eligible students do not apply for scholarships. But although scholarship recipients are likely to differ from other students in unobservable characteristics, it is unlikely that these characteristics differ to a great extent between full and partial scholarship recipients (which mainly differ regarding their income level). As a final robustness test, I estimate the differential effect of receiving a full *Prouni* scholarship relative to a partial *Prouni* scholarship, and then compare it to the difference in coefficients obtained for the two types of scholarship when performing separate estimations. If previous estimates are biased by unobserved student characteristics, it is likely that estimating differential effects will yield different results, as in this case unobserved characteristics will be balanced between both groups and uncorrelated with treatment status.

The results are shown in Tables 1.12 and 1.13. For each field of study, the first line shows the differential effect of receiving a full *Prouni* scholarship relative to a partial *Prouni* scholarship. The second line shows the difference between previously estimated effects of receiving a full and partial *Prouni* scholarship relative to students that received no scholarship. The two types of estimates are very close, although in the case of student performance the differential effects are slightly lower than the difference between the effect of receiving a full and a partial scholarship.

1.6 Concluding Remarks

As higher education becomes increasingly important in the labor market, student financial aid in the form of grants or scholarships is likely to have a key role in increasing access and improving college outcomes for disadvantaged students. However, evidence on the effectiveness of student aid programs is limited, especially in the context of developing countries.

This paper uses propensity score matching to estimate the effects of a scholarship program for disadvantaged college students in Brazil, the *Prouni*, which requires students to pass at least 75% of the classes taken in a given term. By creating a counterfactual of students with similar observable characteristics, I show that final-year students who received a scholarship at some point during their studies perform better in a test measuring skills specific to their field of study, and take less time to reach their final year of college. Those who received a full scholarship also have a lower probability to be working while in college, although this is not observed for partial scholarship recipients. Both types of scholarship recipients also report studying more on average.

These results should be interpreted with caution, as their validity rests on the relatively strong hypothesis of selection on observables. The possibility that scholarship recipients have unobserved characteristics that are both correlated with the probability of treatment and with students' outcomes cannot be completely excluded. However, the fact that the estimated effects are higher among those who received a full scholarship than among those who received a partial scholarship provides some evidence in favor of the validity of the results, as unobserved characteristics are unlikely to differ strongly between both types of students. An interesting direction for future research would be to estimate the impacts of the *Prouni* program on the decision of eligible students to pursue higher education, and on student dropout.

References

Angrist, J., Autor, D., Hudson, S., and Pallais, A. (2014). Leveling Up: Early Results from a Randomized Evaluation of Post-Secondary Aid. *National Bureau of Economic Research Working Paper Series*, No. 20800.

Binelli, C., Meghir, C., and Menezes-Filho, N. (2008). Education and Wages in Brazil. *Institute for Fiscal Studies, mimeo.*

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1):31–72.

Canton, E. and Blom, A. (2004). Can Student Loans Improve Accessibility to Higher Education and Student Performance? *World Bank Policy Research Working Paper*, 3425.

Cornwell, C., Mustard, D. B., and Sridhar, D. J. (2006). The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Program. *Journal of Labor Economics*, 24(4).

Dearden, L., Fitzsimons, E., and Wyness, G. (2014). Money for nothing: Estimating the impact of student aid on participation in higher education. *Economics of Education Review*, 43:66–78.

Dynarski, S. M. (2003). Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *The American Economic Review*, 93(1):279–288.

Kane, T. J. (2003). A Quasi-Experimental Estimate of the Impact of Financial Aid on College-Going. *National Bureau of Economic Research Working Paper Series*, No. 9703.

Murakami, Y. and Blom, A. (2008). Accessibility and Affordability of Tertiary Education in Brazil, Colombia, Mexico and Peru within a Global Context. *The World Bank Policy Research Working Paper Series*, No. 4517.

Nielsen, H. S., Sørensen, T., and Taber, C. R. (2008). Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform. *National Bureau of Economic Research Working Paper Series*, No. 14535.

Rosenbaum, P. R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1):41–55.

Sianesi, B. (2004). An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. *Review of Economics and Statistics*, 86(1):133–155.

Tables and Figures

Table 1.1: Fields of study tested in *Enade* in 2004-2010

2004	Veterinary Medicine, Dentistry, Medicine, Agronomy, Pharmacy, Nursing, Speech Therapy, Nutrition, Physical Education, Physiotherapy, Social Work, Zootechnics, Occupational Therapy
2005	Math, Literature, Physics, Chemistry, Biology, Pedagogy, Architecture, History, Geography, Philosophy, Computer Science, Engineering, Social Sciences
2006	Management, Law, Communication Studies, Economics, Psychology, Accounting, Design, Tourism, Acting, Music, Biomedical Science, Archiving, Library Science, Administrative Assistant Degree, Teaching
2007	Same as 2004 + Biomedical Science, technical degrees in Radiology and Agro-industry
2008	Same as 2005 + technical degrees in Food Science, Systems Analysis and Development, Industrial Automation, Building, Manufacturing Engineering, Production Management, Industrial Maintenance, Chemical Processes, Computer Networks, Environmental Sanitation
2009	Same as 2006 (except Biomedical Science) + International Relations, Statistics, and technical degrees in Design, Marketing, Management, Human Resources, Financial Management, Gastronomy, Tourism
2010	Same as 2007 + technical degrees in Agribusiness, Hospital Management, Environmental Management

Table 1.2: Descriptive statistics - sample of final-year students

	Management	Law	Accounting
No. colleges	581	241	152
No. test-takers	67,450	49,529	16,803
Avg. test-takers per college	114	213	109
% full Prouni scholarship	9	8	8
% partial Prouni scholarship	4	3	5
% other scholarships	2	2	2
% female	55	52	58
% black	6	6	7
Avg. age	27	29	28
% HH income $<$ 3 min. wages	25	22	28
% studied in public high school	62	40	70
% mother finished high school	45	54	38
% father finished high school	43	53	35
% evening study	89	68	95
% working	83	59	89
Avg. years since start of college	3.5	4.3	3.4
Avg. grade - general knowledge	44	48	39
Avg. grade - specific knowledge	37	52	32

Notes: Grades are given in a scale of 0-100. The minimum wage in 2009 was 465 BRL, equivalent to 230 USD at the time approximately.

Table 1.3: Characteristics of test-takers by scholarship status and type of student

	Full Prouni		Partial Prouni		No scholarship	
	First-year	Final-year	First-year	Final-year	First-year	Final-year
			Management			
No. students	6,194	6,068	3,226	2,038	54,152	43,506
% female	58	59	61	60	56	54
% black	11	14	9	8	6	5
Avg. age	23	25	23	26	24	27
% HH income < 3 min. wages	68	49	56	38	36	20
% studied in public high school	93	93	88	86	60	54
% mother finished high school	36	37	39	37	47	48
% father finished high school	31	32	34	35	44	46
% evening study	84	90	82	89	83	88
% working	64	77	72	84	73	83
Avg. years since start of college	-	3.1	-	3.2	-	3.6
Avg. grade - gen. knowledge	51	53	44	47	38	43
Avg. grade - spec. knowledge	37	46	31	39	28	26
			Law			
No. students	4,504	3,320	1,938	1,274	48,704	34,941
% female	50	56	53	53	55	52
% black	13	16	12	12	6	5
Avg. age	24	26	24	27	25	29
% HH income < 3 min. wages	65	59	53	41	25	17
% studied in public high school	89	90	81	77	36	33
% mother finished high school	43	43	45	46	57	57
% father finished high school	38	39	41	43	56	56
% evening study	66	72	65	70	62	67
% working	53	50	59	59	51	59
Avg. years since start of college	-	3.9	-	4.0	-	4.3
Avg. grade - gen. knowledge	55	55	47	49	45	48
Avg. grade - spec. knowledge	51	58	44	51	42	51
			Accounting			
No. students	1,796	1,298	1,236	559	15,218	10,951
% female	59	56	63	57	60	57
% black	12	14	8	11	7	6
Avg. age	24	26	23	27	25	28
% HH income < 3 min. wages	70	48	60	37	41	24
% studied in public high school	93	95	92	88	72	64
% mother finished high school	34	32	37	34	40	40
% father finished high school	27	27	29	30	36	38
% evening study	94	96	92	92	92	95
% working	68	84	77	87	81	89
Avg. years since start of college	-	3.1	-	3.2	-	3.5
Avg. grade - gen. knowledge	49	47	42	42	36	38
Avg. grade - spec. knowledge	28	40	25	35	23	31

Notes: Grades are given in a scale of 0-100. The minimum wage in 2009 was 465 BRL, equivalent to 230 USD at the time approximately.

Table 1.4: Differences between first-year and final-year students' test scores

		Test scores - general knowledge test		Test scores - specific knowledge test	
		Simple differences	OLS	Simple differences	OLS
Management	Coeff.	0.26***	0.22***	0.64***	0.59***
		(0.00)	(0.00)	(0.00)	(0.00)
	Obs.	200,522	150,713	200,522	150,713
Law	Coeff.	0.14***	0.12***	0.57***	0.56***
		(0.01)	(0.01)	(0.01)	(0.01)
	Obs.	158,289	117,080	158,289	117,080
Accounting	Coeff.	0.12***	0.08***	0.77***	0.73***
		(0.01)	(0.01)	(0.01)	(0.01)
	Obs.	53,355	40,754	53,355	40,754

Note: Outcomes are standardized test scores. In each field, the sample is composed by first-year and final-year students pooled together. The coefficients measure the effect of being a final-year student on test scores, obtained by including a dummy=1 for final-year students. Other OLS controls include: gender (a dummy=1 for females), race (a dummy=1 for black students) mother and father education (a dummy=1 if the mother/father have completed high school), type of high school attended (a dummy=1 if attended a public high school only), household income (dummies for 3 out of 4 categories of income).
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 1.5: Probit estimation of propensity scores

	Management		Law		Accounting	
	Full Prouni	Partial Prouni	Full Prouni	Partial Prouni	Full Prouni	Partial Prouni
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.07***	-0.02	-0.05*	-0.09***	-0.13***	-0.08*
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)
Black	-0.04***	-0.03***	-0.05***	-0.04***	-0.04***	-0.02***
	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Age	0.59***	0.22***	0.59***	0.38***	0.49***	0.27***
	(0.03)	(0.04)	(0.04)	(0.05)	(0.06)	(0.08)
Mother w/ high school	-0.04*	-0.07***	-0.11***	-0.08**	-0.05	-0.04
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)
Father w/ high school	-0.13***	-0.07***	-0.17***	-0.11***	-0.11***	-0.06
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)
Public high school	1.03***	0.60***	0.87***	0.80***	0.89***	0.52***
	(0.03)	(0.03)	(0.07)	(0.03)	(0.06)	(0.06)
HH income=3-6 min. wages	-0.44***	-0.22***	-0.54***	-0.20***	-0.35***	-0.16***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)
HH income=6-10 min. wages	-0.98***	-0.43***	-1.14***	-0.56***	-0.83***	-0.40***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.06)	(0.07)
HH income >10 min. wages	-1.43***	-0.75***	-1.81***	-0.94***	-1.29***	-0.71***
	(0.04)	(0.05)	(0.06)	(0.05)	(0.01)	(0.01)
Evening classes	0.02***	0.02***	0.01***	0.00***	0.01***	0.01***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Other HH members: 1-3	-0.05*	-0.08**	-0.10***	-0.11***	-0.02	-0.31***
	(0.03)	(0.04)	(0.03)	(0.03)	(0.08)	(0.08)
Other HH members: >4	0.06*	0.06	0.15***	0.19***	0.07	0.06
	(0.03)	(0.05)	(0.04)	(0.05)	(0.07)	(0.08)
Gen. knowledge score	0.23***	0.10**	0.39***	0.26***	0.23***	0.16*
	(0.04)	(0.05)	(0.04)	(0.06)	(0.07)	(0.09)
Obs.	47,987	44,102	36,738	34,767	11,822	11,115
Pseudo R-squared	0.2540	0.1016	0.3936	0.1825	0.1920	0.0775

Note: In columns 1, 3, 5 the outcome is a dummy which equals 1 if the student has received a full Prouni scholarship, and 0 if the student has received no scholarship. In columns 2, 4, 6 the outcome is a dummy which equals 1 if the student has received a partial Prouni scholarship, and 0 if the student has received no scholarship. Controls include: gender (a dummy=1 for females), race (a dummy=1 for black students), age, mother and father education (a dummy=1 if the mother/father have completed high school), type of high school attended (a dummy=1 if attended a public high school), income (dummies for 3 out of 4 categories: 0-3 minimum wages, 3-6 minimum wages, 6-10 minimum wages, >10 minimum wages), period of study (a dummy=1 if enrolled in evening classes), household size (dummies for 2 out of 3 categories: 0 other members, 1-3 other members, >4 other members) and general knowledge grade. Interactions and/or quadratics are included in columns 2 and 3 in order to achieve balance. Standard errors in parentheses.

$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 1.6: Results - performance on specific knowledge test

	Simple differences	OLS	PSM Nearest neighbor	PSM 5 nearest neighbors	PSM Radius	PSM Kernel
	(1)	(2)	(3)	(4)	(5)	(6)
			Management			
Full Prouni	0.66***	0.61***	0.63***	0.64***	0.65***	0.65***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Treated	6,068	5,857	5,855	5,855	5,846	5,846
Control	43,506	42,130	42,130	42,130	42,130	42,130
Partial Prouni	0.19***	0.23***	0.24***	0.23***	0.25***	0.25***
	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)
Treated	2,038	1,972	1,972	1,972	1,972	1,972
Control	43,506	42,130	42,130	42,130	42,130	42,130
			Law			
Full Prouni	0.39***	0.36***	0.37***	0.38***	0.38***	0.37***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Treated	3,320	3,180	3,172	3,172	3,159	3,159
Control	34,941	33,558	33,558	33,558	33,558	33,558
Partial Prouni	0.01	0.08**	0.08**	0.12***	0.11***	0.11***
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Treated	1,274	1,209	1,209	1,209	1,209	1,209
Control	34,941	33,558	33,558	33,558	33,558	33,558
			Accounting			
Full Prouni	0.62***	0.59***	0.63***	0.63***	0.62***	0.62***
	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Treated	1,298	1,242	1,236	1,236	1,236	1,236
Control	10,951	10,580	10,580	10,580	10,580	10,580
Partial Prouni	0.25***	0.27***	0.33***	0.28***	0.29***	0.29***
	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)
Treated	559	535	535	535	535	535
Control	10,951	10,580	10,580	10,580	10,580	10,580

Note: Outcomes are standardized test scores. OLS controls include the same variables used in propensity score estimations. Standard errors in parentheses (OLS standard errors clustered at the college level). In columns 3 to 5, matching is done with replacement. In column 5, a caliper of 0.01 is used. In column 6, the Epanechnikov kernel function is used and a bandwidth of 0.01.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 1.7: Results - duration of studies

	Simple differences	OLS	PSM Nearest neighbor	PSM 5 nearest neighbors	PSM Radius	PSM Kernel
	(1)	(2)	(3)	(4)	(5)	(6)
			Management			
Full Prouni	-0.47***	-0.22***	-0.23***	-0.23***	-0.22***	-0.22***
	(0.05)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Treated	6,068	5,857	5,855	5,855	5,846	5,846
Control	43,506	42,130	42,130	42,130	42,130	42,130
Partial Prouni	-0.34***	-0.14***	-0.19***	-0.14***	-0.15***	-0.15***
	(0.06)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
Treated	2,038	1,972	1,972	1,972	1,972	1,972
Control	43,506	42,130	42,130	42,130	42,130	42,130
			Law			
Full Prouni	-0.43***	-0.34***	-0.30***	-0.33***	-0.34***	-0.34***
	(0.06)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
Treated	3,320	3,180	3,172	3,172	3,159	3,172
Control	34,941	33,558	33,558	33,558	33,558	33,558
Partial Prouni	-0.31***	-0.26***	-0.24***	-0.25***	-0.26***	-0.26***
	(0.05)	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)
Treated	1,274	1,209	1,209	1,209	1,209	1,209
Control	34,941	33,558	33,558	33,558	33,558	33,558
			Accounting			
Full Prouni	-0.36***	-0.24***	-0.24***	-0.20***	-0.21***	-0.21***
	(0.07)	(0.06)	(0.06)	(0.03)	(0.04)	(0.04)
Treated	1,298	1,242	1,236	1,236	1,236	1,236
Control	10,951	10,580	10,580	10,580	10,580	10,580
Partial Prouni	-0.29***	-0.20**	-0.14*	-0.18***	-0.20***	-0.20***
	(0.09)	(0.09)	(0.08)	(0.06)	(0.06)	(0.06)
Treated	559	535	535	535	535	535
Control	10,951	10,580	10,580	10,580	10,580	10,580

Note: Outcome is the duration of studies, measured by the number of years since enrolling in college. OLS controls include the same variables used in propensity score estimations. Standard errors in parentheses (OLS standard errors clustered at the college level). In columns 3 to 5, matching is done with replacement. In column 5, a caliper of 0.01 is used. In column 6, the Epanechnikov kernel function is used and a bandwidth of 0.01.
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 1.8: Results - work while in college

	PSM Nearest neighbor (1)	PSM 5 nearest neighbors (2)	PSM Radius (3)	PSM Kernel (4)
		Management		
Full Prouni	-0.03***	-0.04***	-0.03***	-0.03***
	(0.01)	(0.01)	(0.01)	(0.01)
Treated	5,825	5,826	5,817	5,817
Control	41,921	41,921	41,921	41,921
Partial Prouni	0.03**	0.02*	0.02**	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)
Treated	1,958	1,958	1,957	1,957
Control	41,921	41,921	41,921	41,921
		Law		
Full Prouni	-0.07***	-0.07***	-0.07***	-0.07***
	(0.01)	(0.01)	(0.01)	(0.01)
Treated	3,154	3,154	3,140	3,140
Control	33,378	33,378	33,378	33,378
Partial Prouni	0.00	0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
Treated	1,199	1,199	1,199	1,199
Control	33,378	33,378	33,378	33,378
		Accounting		
Full Prouni	-0.06***	-0.05***	-0.04***	-0.05***
	(0.02)	(0.01)	(0.02)	(0.01)
Treated	1,227	1,227	1,226	1,226
Control	10,516	10,516	10,516	10,516
Partial Prouni	-0.01	-0.03	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)
Treated	532	532	532	532
Control	10,516	10,516	10,516	10,516

Note: The outcome is a dummy=1 if the student is working. Standard errors in parentheses. In columns 3 to 5, matching is done with replacement. In column 5, a caliper of 0.01 is used. In column 6, the Epanechnikov kernel function is used and a bandwidth of 0.01.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 1.9: Results - time spent studying

	PSM Nearest neighbor (1)	PSM 5 nearest neighbors (2)	PSM Radius (3)	PSM Kernel (4)
		Management		
Full Prouni	0.08***	0.09***	0.09***	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)
Treated	5,825	5,826	5,817	5,817
Control	41,921	41,921	41,921	41,921
Partial Prouni	0.06***	0.05***	0.05***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)
Treated	1,958	1,958	1,957	1,957
Control	41,921	41,921	41,921	41,921
		Law		
Full Prouni	0.10***	0.10***	0.10***	0.10***
	(0.02)	(0.01)	(0.01)	(0.01)
Treated	3,154	3,154	3,140	3,140
Control	33,378	33,378	33,378	33,378
Partial Prouni	0.05**	0.05***	0.05***	0.05***
	(0.02)	(0.02)	(0.02)	(0.02)
Treated	1,199	1,199	1,199	1,199
Control	33,378	33,378	33,378	33,378
		Accounting		
Full Prouni	0.04*	0.06***	0.05***	0.05***
	(0.02)	(0.02)	(0.02)	(0.02)
Treated	1,227	1,227	1,226	1,226
Control	10,516	10,516	10,516	10,516
Partial Prouni	0.02	0.03	0.04**	0.04**
	(0.03)	(0.02)	(0.02)	(0.02)
Treated	532	532	532	532
Control	10,516	10,516	10,516	10,516

Note: The outcome is a dummy=1 if the student reports studying more than 3 hours a week. In columns 3 to 5, matching is done with replacement. In column 5, a caliper of 0.01 is used. In column 6, the Epanechnikov kernel function is used and a bandwidth of 0.01.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 1.10: Robustness - performance on specific knowledge test with college controls

	OLS w/ college controls (1)	OLS w/ college dummies (2)	PSM Nearest neighbor (3)	PSM 5 nearest neighbors (4)	PSM Radius (5)	PSM Kernel (6)
			Management			
Full Prouni	0.53***	0.56***	0.57***	0.58***	0.57***	0.57***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Treated	5,780	5,857	5,776	5,775	5,764	5,764
Control	41,793	42,130	41,793	41,793	41,793	41,793
Partial Prouni	0.21***	0.23***	0.19***	0.23***	0.23***	0.23***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Treated	1,936	1,972	1,936	1,936	1,936	1,936
Control	41,793	42,130	41,793	41,793	41,793	41,793
			Law			
Full Prouni	0.33***	0.34***	0.35***	0.34***	0.35***	0.35***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Treated	3,161	3,180	3,159	3,159	3,137	3,137
Control	33,438	33,558	33,438	33,438	33,438	33,438
Partial Prouni	0.09***	0.10***	0.13***	0.11***	0.11***	0.12***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Treated	1,201	1,209	1,201	1,201	1,201	1,201
Control	33,438	33,558	33,438	33,438	33,438	33,438
			Accounting			
Full Prouni	0.54***	0.58***	0.55***	0.57***	0.56***	0.56***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Treated	1,187	1,242	1,187	1,187	1,181	1,181
Control	10,170	10,580	10,170	10,170	10,170	10,170
Partial Prouni	0.25***	0.29***	0.29***	0.28***	0.28***	0.27***
	(0.04)	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)
Treated	497	535	497	497	497	497
Control	10,170	10,580	10,170	10,170	10,170	10,170

Note: Outcomes are standardized test scores. OLS controls in column 1 include the same variables used in previous propensity score estimations, as well as average test scores of first-year and final-year students at the general and specific knowledge tests (from students in the same field and college). OLS controls in column 2 include the same variables used in previous propensity score estimations, as well as college dummies. Standard errors in parentheses (OLS standard errors clustered at the college level). Propensity scores are estimated using the same variables used as OLS controls in column 1. In columns 3 to 5, matching is done with replacement. In column 5, a caliper of 0.01 is used. In column 6, the Epanechnikov kernel function is used and a bandwidth of 0.01.

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 1.11: Robustness - duration of studies with college controls

	OLS w/ college controls	OLS w/ college dummies	PSM Nearest neighbor	PSM 5 nearest neighbors	PSM Radius	PSM Kernel
	(1)	(2)	(3)	(4)	(5)	(6)
			Management			
Full Prouni	-0.28***	-0.28***	-0.28***	-0.29***	-0.27***	-0.27***
	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)
Treated	5,780	5,857	5,776	5,775	5,764	5,764
Control	41,793	42,130	41,793	41,793	41,793	41,793
Partial Prouni	-0.15***	-0.12**	-0.16**	-0.16**	-0.16**	-0.16**
	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
Treated	1,936	1,972	1,936	1,936	1,936	1,936
Control	41,793	42,130	41,793	41,793	41,793	41,793
			Law			
Full Prouni	-0.34***	-0.29***	-0.32***	-0.32***	-0.34***	-0.34***
	(0.05)	(0.07)	(0.04)	(0.03)	(0.03)	(0.03)
Treated	3,161	3,180	3,159	3,159	3,137	3,137
Control	33,438	33,558	33,438	33,438	33,438	33,438
Partial Prouni	-0.28***	-0.13***	-0.28***	-0.28***	-0.28***	-0.28***
	(0.05)	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)
Treated	1,201	1,209	1,201	1,201	1,201	1,201
Control	33,438	33,558	33,438	33,438	33,438	33,438
			Accounting			
Full Prouni	-0.26***	-0.27***	-0.22***	-0.25***	-0.24***	-0.25***
	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
Treated	1,187	1,242	1,187	1,187	1,181	1,181
Control	10,170	10,580	10,170	10,170	10,170	10,170
Partial Prouni	-0.19**	-0.13*	-0.20**	-0.19***	-0.18***	-0.18***
	(0.09)	(0.07)	(0.09)	(0.07)	(0.06)	(0.06)
Treated	497	535	497	497	497	497
Control	10,170	10,580	10,170	10,170	10,170	10,170

Note: Outcome is the duration of studies, measured by the number of years since enrolling in college. OLS controls in column 1 include the same variables used in previous propensity score estimations, as well as average test scores of first-year and final-year students at the general and specific knowledge tests (from students in the same field and college). OLS controls in column 2 include the same variables used in previous propensity score estimations, as well as college dummies. Standard errors in parentheses (OLS standard errors clustered at the college level). Propensity scores are estimated using the same variables used as OLS controls in column 1. In columns 3 to 5, matching is done with replacement. In column 5, a caliper of 0.01 is used. In column 6, the Epanechnikov kernel function is used and a bandwidth of 0.01.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 1.12: Differential effects - performance on specific knowledge test

	Simple differences	OLS	PSM Nearest neighbor	PSM 5 nearest neighbors	PSM Radius	PSM Kernel
	(1)	(2)	(3)	(4)	(5)	(6)
			Management			
Differential effect	0.46***	0.37***	0.33***	0.36***	0.37***	0.37***
	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Diff. in coefficients	0.47	0.38	0.39	0.41	0.40	0.40
			Law			
Differential effect	0.38***	0.25***	0.20***	0.21***	0.22***	0.22***
	(0.03)	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)
Diff. in coefficients	0.38	0.28	0.29	0.26	0.27	0.26
			Accounting			
Differential effect	0.37***	0.31***	0.27***	0.32***	0.32***	0.32***
	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)
Diff. in coefficients	0.37	0.32	0.30	0.35	0.33	0.33

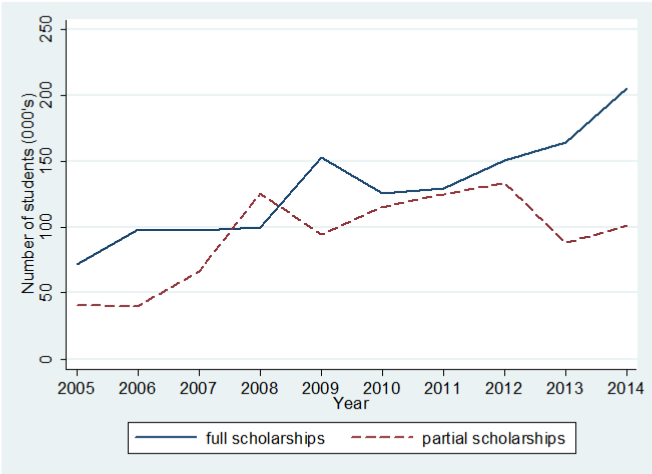
Note: Outcomes are standardized test scores. The first line shows the differential effect of receiving a full Prouni scholarship relative to a partial Prouni scholarship (coefficients are obtained by matching full Prouni recipients to partial Prouni recipients). The second line shows the difference between previously estimated coefficients for full and partial Prouni students separately (obtained by matching full and partial Prouni recipients to students without scholarships).
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 1.13: Differential effects - duration of studies

	Simple differences	OLS	PSM Nearest neighbor	PSM 5 nearest neighbors	PSM Radius	PSM Kernel
	(1)	(2)	(3)	(4)	(5)	(6)
			Management			
Differential effect	-0.12***	-0.09**	-0.02	-0.05*	-0.05*	-0.05*
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Diff. in coefficients	-0.13	-0.08	-0.04	-0.09	-0.07	-0.07
			Law			
Differential effect	-0.12***	-0.10**	-0.18***	-0.12***	-0.10***	-0.10***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
Diff. in coefficients	-0.12	-0.08	-0.06	-0.08	-0.08	-0.08
			Accounting			
Differential effect	-0.07	-0.02	0.00	0.01	-0.01	-0.01
	(0.07)	(0.06)	(0.09)	(0.07)	(0.07)	(0.07)
Diff. in coefficients	-0.07	-0.04	-0.10	-0.02	-0.01	-0.01

Note: Outcome is the duration of studies, measured by the number of years since enrolling in college. The first line shows the differential effect of receiving a full Prouni scholarship relative to a partial Prouni scholarship (coefficients are obtained by matching full Prouni recipients to partial Prouni recipients). The second line shows the difference between previously estimated coefficients for full and partial Prouni students separately (obtained by matching full and partial Prouni recipients to students without scholarships).
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Figure 1.1: Evolution of Prouni scholarships over time



Source: Ministry of Education

Appendix to Chapter 1

Table 1.A1: Balance in covariates before matching - Management

	Controls (1)	Full Prouni (2)	Partial Prouni (3)	Diff. (1)-(2)	p-value	Diff. (1)-(3)	p-value
% female	54.0	59.0	59.7	-5.0	0.000	-5.7	0.000
Avg. age	27.2	25.3	25.7	1.9	0.000	1.6	0.000
% black	4.7	13.8	7.9	-9.2	0.000	-3.3	0.000
% father finished high school	46.5	31.7	34.5	14.7	0.000	11.9	0.000
% mother finished high school	48.3	36.9	37.3	11.4	0.000	11.0	0.000
% studied in public high school	54.2	93.2	85.8	-39.0	0.000	-31.6	0.000
% HH income 0-3 min. wages	20.4	49.6	38.4	-29.2	0.000	-17.9	0.000
% HH income 3-6 min. wages	34.5	40.0	41.3	-5.5	0.000	-6.8	0.000
% HH income 6-10 min. wages	22.8	8.5	15.5	14.4	0.000	7.4	0.000
% HH income >10 min. wages	22.2	2.0	4.9	20.3	0.000	17.4	0.000
Avg. grade - gen. knowledge	42.8	53.1	47.2	-10.3	0.000	-4.4	0.000
% evening study	88.1	89.7	88.9	-1.5	0.000	-0.8	0.277
Other HH members: 0	6.6	8.1	7.2	-1.5	0.000	-0.6	0.260
Other HH members: 1-3	62.7	57.2	61.3	5.6	0.000	1.5	0.186
Other HH members >4	30.7	34.7	31.5	-4.0	0.000	-0.8	0.436

Table 1.A2: Balance in covariates after matching - Management

	Diff. Controls-Full Prouni	p-value	Diff. Control-Partial Prouni	p-value
% female	0.6	0.486	1.0	0.536
Avg. age	-0.1	0.125	-0.2	0.316
% black	0.1	0.915	-0.9	0.277
% father finished high school	-0.1	0.952	-1.4	0.347
% mother finished high school	-0.3	0.730	-1.0	0.530
% studied in public high school	0.3	0.577	-0.5	0.681
% HH income 0-3 min. wages	-0.2	0.868	0.2	0.896
% HH income 3-6 min. wages	0.6	0.474	0.0	0.974
% HH income 6-10 min. wages	-0.6	0.252	0.2	0.895
% HH income >10 min. wages	0.1	0.743	-0.3	0.652
Avg. grade - gen. knowledge	-0.5	0.151	-0.4	0.562
% evening study	0.8	0.137	1.8	0.056
Other HH members: 0	-0.9	0.059	-0.6	0.490
Other HH members: 1-3	1.6	0.085	0.8	0.623
Other HH members >4	-0.6	0.460	-0.2	0.891

Note: Nearest neighbor propensity score matching with replacement is used

Table 1.A3: Balance in covariates before matching - Law

	Controls (1)	Full Prouni (2)	Partial Prouni (3)	Diff. (1)-(2)	p-value	Diff. (1)-(3)	p-value
% female	51.7	56.3	53.2	-4.6	0.000	-1.6	0.270
Avg. age	29.0	25.7	26.7	3.4	0.000	2.3	0.000
% black	4.6	15.7	12.3	-11.0	0.000	-7.7	0.000
% father finished high school	55.9	38.7	42.6	17.2	0.000	13.3	0.000
% mother finished high school	56.5	42.6	45.9	13.9	0.000	10.7	0.000
% studied in public high school	32.5	89.5	77.3	-57.0	0.000	-44.8	0.000
% HH income 0-3 min. wages	16.8	59.4	41.9	-42.6	0.000	-25.1	0.000
% HH income 3-6 min. wages	24.7	32.0	39.1	-7.3	0.000	-14.4	0.000
% HH income 6-10 min. wages	22.2	7.0	13.1	15.1	0.000	9.1	0.000
% HH income >10 min. wages	36.3	1.6	6.1	34.8	0.000	30.3	0.000
Avg. grade - gen. knowledge	47.7	54.7	48.6	-7.0	0.000	-0.9	0.132
% evening study	66.6	71.8	69.8	-5.3	0.000	-3.3	0.016
Other HH members: 0	8.6	10.3	8.7	-1.8	0.001	-0.1	0.863
Other HH members: 1-3	62.3	54.8	59.9	7.5	0.000	2.4	0.088
Other HH members >4	29.2	34.9	31.4	-5.7	0.000	-2.2	0.087

Table 1.A4: Balance in covariates after matching - Law

	Diff. Controls-Full Prouni	p-value	Diff. Control-Partial Prouni	p-value
% female	0.6	0.649	0.6	0.775
Avg. age	0.2	0.121	0.3	0.195
% black	0.9	0.354	0.0	1.000
% father finished high school	-1.4	0.245	0.3	0.869
% mother finished high school	-0.3	0.800	-0.3	0.870
% studied in public high school	0.1	0.902	1.9	0.253
% HH income 0-3 min. wages	1.4	0.271	-0.7	0.741
% HH income 3-6 min. wages	-1.7	0.144	0.8	0.677
% HH income 6-10 min. wages	-0.3	0.693	-1.0	0.467
% HH income >10 min. wages	0.6	0.074	0.8	0.408
Avg. grade - gen. knowledge	0.2	0.763	0.1	0.885
% evening study	1.6	0.143	1.4	0.448
Other HH members: 0	0.3	0.682	-1.2	0.295
Other HH members: 1-3	-1.5	0.217	0.8	0.677
Other HH members >4	1.2	0.306	0.3	0.861

Note: Nearest neighbor propensity score matching with replacement is used

Table 1.A5: Balance in covariates before matching - Accounting

	Controls (1)	Full Prouni (2)	Partial Prouni (3)	Diff. (1)-(2)	p-value	Diff. (1)-(3)	p-value
% female	57.4	56.4	56.7	1.0	0.500	0.7	0.756
Avg. age	28.5	25.7	26.7	2.8	0.000	1.8	0.000
% black	5.8	14.4	10.8	-8.6	0.000	-5.0	0.000
% father finished high school	37.6	27.1	30.0	10.5	0.000	7.6	0.000
% mother finished high school	39.7	32.3	34.2	7.5	0.000	5.5	0.009
% studied in public high school	64.4	94.7	88.0	-30.3	0.000	-23.6	0.000
% HH income 0-3 min. wages	24.1	48.4	37.9	-24.4	0.000	-13.9	0.000
% HH income 3-6 min. wages	38.0	41.6	44.1	-3.6	0.012	-6.1	0.004
% HH income 6-10 min. wages	22.8	8.6	14.3	14.2	0.000	8.5	0.000
% HH income >10 min. wages	15.1	1.4	3.6	13.7	0.000	11.5	0.000
Avg. grade - gen. knowledge	38.2	47.3	42.5	-9.2	0.000	-4.3	0.000
% evening study	94.7	95.7	91.9	-1.0	0.127	2.8	0.004
Other HH members: 0	7.2	8.6	7.4	-1.4	0.064	-0.2	0.874
Other HH members: 1-3	61.5	56.7	57.4	4.8	0.001	4.1	0.055
Other HH members >4	31.4	34.7	35.3	-3.4	0.015	-3.9	0.054

Table 1.A6: Balance in covariates after matching - Accounting

	Diff. Controls-Full Prouni	p-value	Diff. Control-Partial Prouni	p-value
% female	1.9	0.329	3.6	0.238
Avg. age	0.0	0.844	0.1	0.770
% black	0.2	0.863	-1.9	0.298
% father finished high school	-1.9	0.273	-3.2	0.251
% mother finished high school	-0.2	0.931	-1.1	0.699
% studied in public high school	-0.2	0.795	2.6	0.166
% HH income 0-3 min. wages	1.7	0.398	1.3	0.661
% HH income 3-6 min. wages	-1.4	0.487	-0.7	0.806
% HH income 6-10 min. wages	-0.8	0.463	-1.1	0.594
% HH income >10 min. wages	0.5	0.339	0.6	0.641
Avg. grade - gen. knowledge	-0.5	0.525	-0.2	0.866
% evening study	0.5	0.544	1.1	0.491
Other HH members: 0	-0.5	0.664	-0.2	0.907
Other HH members: 1-3	-2.2	0.274	3.2	0.290
Other HH members >4	2.7	0.165	-3.0	0.300

Note: Nearest neighbor propensity score matching with replacement is used

Table 1.A7: Pseudo R-squared before and after matching

	Before matching	**After matching**			
		Nearest neighbor	5 nearest neighbors	Radius	Kernel
			Management		
Full Prouni	0.254	0.0009	0.0009	0.0006	0.0005
Partial Prouni	0.1016	0.002	0.0012	0.0004	0.0004
			Law		
Full Prouni	0.3936	0.0016	0.002	0.0032	0.003
Partial Prouni	0.1825	0.002	0.0018	0.0025	0.0024
			Accounting		
Full Prouni	0.192	0.003	0.0033	0.0036	0.0035
Partial Prouni	0.0775	0.002	0.0018	0.0025	0.0024

Note: McFadden's pseudo R-squared are presented, from probit estimations where the probability of treatment is estimated separately for full and partial Prouni scholarships.

Figure 1.A1: Kernel density estimates of the distribution of propensity scores for Management students (nearest neighbor matching with replacement)

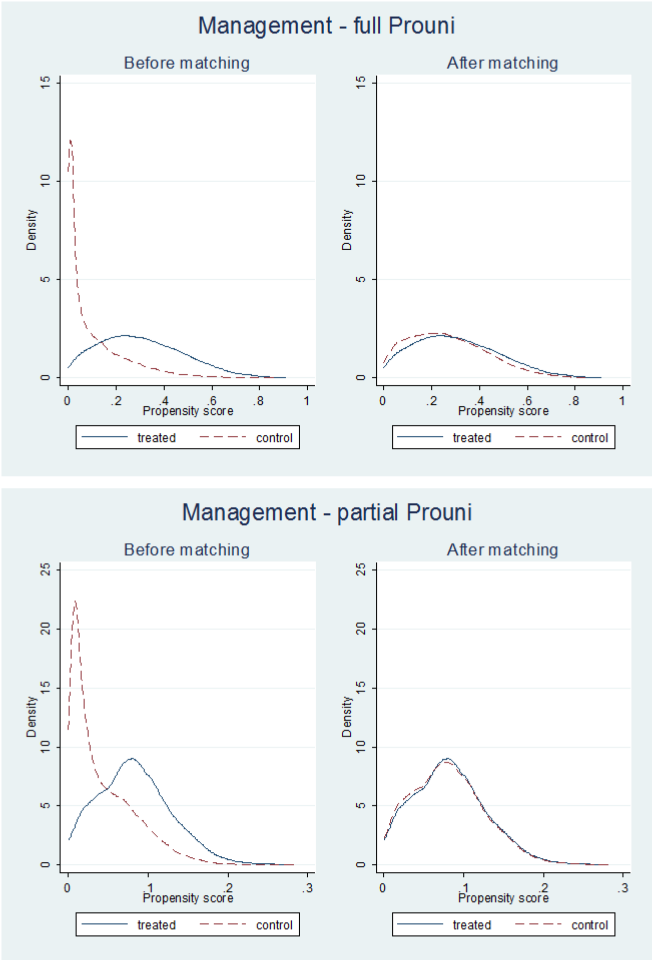


Figure 1.A2: Kernel density estimates of the distribution of propensity scores for Law students (nearest neighbor matching with replacement)

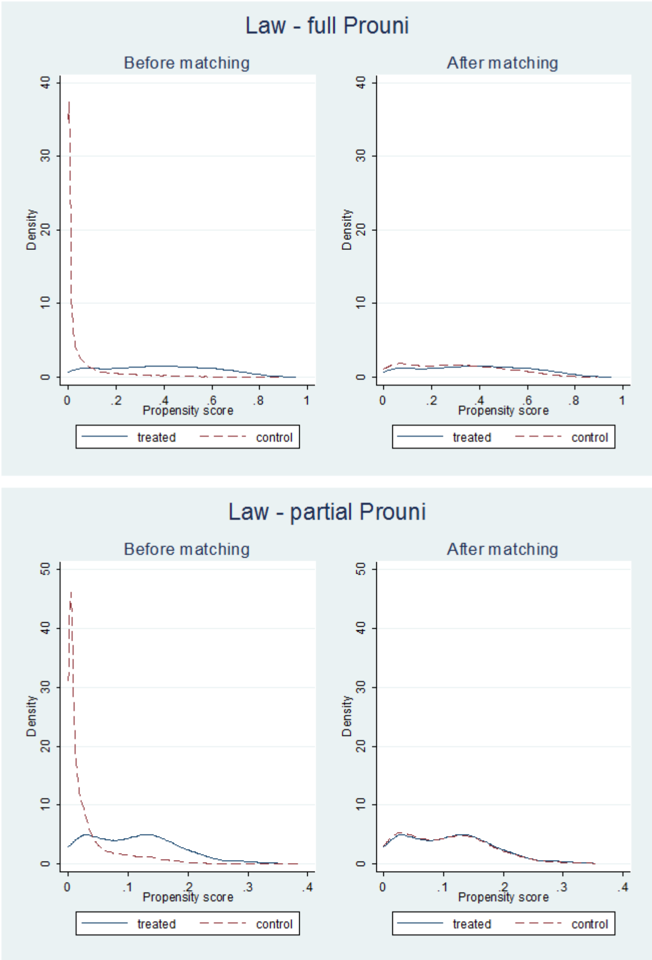
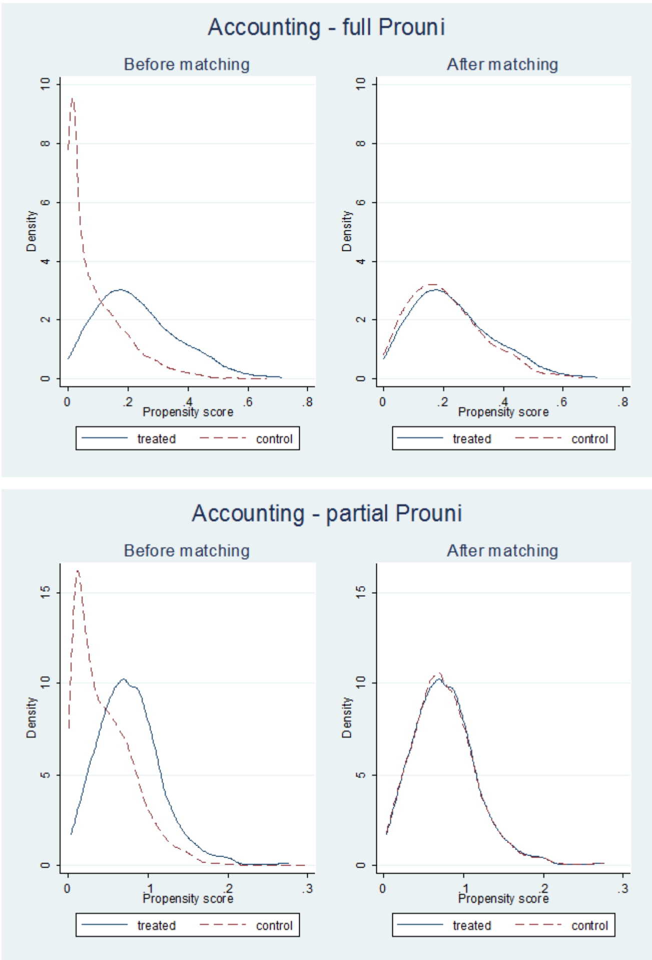


Figure 1.A3: Kernel density estimates of the distribution of propensity scores for Accounting students (nearest neighbor matching with replacement)



2 Teacher Incentives and Student Performance: Evidence from Brazil

2.1 Introduction

Policies linking teacher pay to student performance have drawn considerable attention in recent years, and are seen as a promising way of improving student learning. This idea is motivated by the view that teachers generally face weak incentives and low accountability, especially in developing countries. And while teacher quality has been shown to be an important factor explaining student achievement (Rockoff 2004), teacher observable characteristics such as experience and qualifications, which are the main determinants of salaries in most school systems, do not seem to be good predictors of student performance (Rivkin, Hanushek and Kain 2005; Aaronson, Barrow and Sander 2007).

It is not obvious, however, that incentives will have the desired effects in the specific context of schools. While the standard principal-agent framework suggests that appropriate incentives can increase efficiency, schools have specific characteristics that could change the classic model's predictions. First, given that teachers' work involves multiple tasks, the introduction of performance pay could result in a reallocation of effort toward skills tested on exams linked to the incentive program (or "teaching to the test"), at the expense of human capital accumulation in a broader sense. Recent papers have discussed this issue (Muralidharan and Sundararaman 2011; Neal 2011) based on the theoretical model developed by Holmstrom and Milgrom (1991), and show that the final outcome depends on model parameters of a given school system, which are generally not known. Second, the fact that teaching involves teamwork means there can be complementarities in the education production function, which explains why teacher performance bonuses are often distributed at the group level. One issue related to this type of incentive design is the possible occurrence of free-riding (Holmstrom 1982), which is likely to increase as the group gets large and may limit the effectiveness of such policies. Third, it is possible that in organizations where agents are driven by intrinsic motivation such as schools, incentives can become less efficient (Besley and Ghatak 2005). Finally, incentives could be inefficient if they are not properly understood by teachers. Overall, the theoretical literature indicates that the impact of this type of program is likely to be context-dependent, pointing to the importance of empirical evidence.

Rigorous empirical evidence on the effectiveness of such policies is limited, and points to mixed findings. Fryer (2013) uses a randomized trial to analyze the effects of a teacher incentive program in New York, where schools could decide how they would distribute the incentive bonus among teachers, and finds no evidence that the program increased student performance or changed teacher behavior. Other studies in the United States (Glazerman and Seifullah 2012; Springer et al. 2010) also find no clear evidence that teacher incentive programs improved student achievement. In England, the impact of a pay scheme for teachers partly based on student performance was estimated by Atkinson et al. (2009). Although they find overall positive effects, the impacts are negative for some subjects. A similar program is analyzed by Martins (2009), who looks at the effect of a teacher pay reform in Portugal which conditioned teachers' progression on a pay scale on student test scores, and finds that the policy resulted in a decline in student performance and an increase in grade inflation.

Studies from developing countries provide more encouraging findings. In a well-known paper, Muralidharan and Sundararaman (2011) present results from a randomized evaluation in India assessing the impact of a program that provided group and individual bonuses for teachers, conditional on student performance. They find that both types of programs increased student achievement, and find no evidence of adverse behavior from teachers. In another randomized trial by Glewwe, Ilias and Kremer (2010) in Kenya, a school-based incentive program directed at teachers from primary schools was found to have positive but short-term effects on students' test scores. However, gains were only observed for exams directly linked to the incentives, showing evidence of "teaching to the test" and suggesting that the program did not lead to a broad increase in learning. Lavy (2002, 2009) looks at the impact of two different teacher incentive programs in Israeli high schools using natural experiments, and also finds positive effects on student test scores.

This paper contributes to the literature on teacher incentives by evaluating the effect of a group-based teacher pay for performance program in the state of São Paulo, Brazil, introduced in 2008. The program awards bonuses to teachers and staff from public schools run by the state government, conditional on improvements in student test scores. Using a difference-in-differences and triple differences framework and data from a standardized test not directly related to the incentive scheme (*Prova Brasil*), I estimate the effect of the program on 5th and 9th grade students up to 5 years after the program implementation. The robustness of the results is assessed through the use of a series of different counterfactuals. The estimates obtained show that the program had overall positive effects, although achievement gains among 9th grade students were more modest and less robust across counterfactuals. When averaging the estimates found across all counterfactuals, 5th grade students' math and language test scores improved by 0.15 and 0.09 standard deviations respectively, while the corresponding gains for 9th grade students were 0.06 and 0.03 standard deviations.

I also investigate whether initial school characteristics affect the impact of the program. First, I look at whether the number of teachers influences how schools react to the program, in an attempt to provide evidence on the presence of possible free-riding effects. This does not seem to be generally the case, although the number of teachers has a modest negative effect on 9th grade students' performance. I also estimate how schools' initial performance affect gains from the program, and find that initially low-performing schools improved much more than other schools, suggesting a reduction in inequality among schools.

A related study by Oshiro, Scorzafave and Dorigan (2015) assesses the impact of the teacher bonus program in São Paulo through propensity score matching and difference-in-differences and finds mixed results. This paper aims at bringing further evidence on this issue by using a different methodology and a longer span of data, and by exploring additional impacts of the policy.

This paper is organized as follows. Section 2.2 provides background information on the teacher performance program in São Paulo. Section 2.3 describes the data used and shows some descriptive statistics. Section 2.4 discusses the empirical strategy, and Section 2.5 presents the main results. Heterogeneous effects are analyzed in Section 2.6, and Section 2.7 concludes.

2.2 Background

São Paulo is the richest and most populous state in Brazil, with a population of over 40 million. As in the rest of Brazil, its school system is composed of private schools, and of public schools managed either by the federal, state or municipal government (from now on referred to as "federal schools", "state schools" and "municipal schools" respectively). According to school census data, the state government managed over 5500 schools in 2013, which accounted for around 20% of the total number of schools in São Paulo[1].

Despite improvements in recent years, student performance in São Paulo and in Brazil more generally remain low compared to high-income countries, and even compared to countries with similar levels of per capita income, as evidenced by international student assessments such as PISA[2]. As in many other developing countries, teachers working in the Brazilian public school system face low accountability and cannot be easily dismissed for poor performance. Teacher absenteeism is also an important problem.

The pay for performance program is part of an initiative launched in 2008 by the Secretary

[1] Private schools and municipal schools accounted for approximately 35% and 44% of schools respectively.

[2] Program for International Student Assessment of the OECD (Organisation for Economic Cooperation and Development).

of Education of the State of São Paulo, with the objective of improving the quality of education in public schools run by the state government. According to this initiative, schools are assessed every year through an indicator called *Idesp* (*Índice de Desenvolvimento da Educação do Estado de São Paulo*), which serves as the basis for the calculation of teacher bonuses. Although discussions on the pay for performance program in São Paulo started in 2007, its practical details were not clear until the end of 2008 and the law determining its adoption was only passed in December 2008. First bonus payments were distributed in March 2009 based on the progress achieved by schools in 2008 relative to 2007 (the first year for which the *Idesp* was calculated).

The *Idesp* indicator combines information on student retention and on schools' performance measured through an annual standardized test, the *Saresp* (*Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo*). Only students' performance in math and language (Portuguese) are taken into account, even though the test covers other subjects such as science. The *Idesp* is calculated separately for three different school cycles: 1st to 5th grade, 6th to 9th grade, and 10th to 12th grade. It is obtained through the multiplication of an index of student performance (which involves a calculation based on the proportion of students that fall in pre-determined performance categories), and an index of student flow, determined by the share of students allowed to pass to the next grade. From 2013 onwards, the *Idesp* also started taking into account students' socio-economic composition.

The attribution of bonuses is linked to schools' *Idesp* improvement relative to targets that are individually set. These targets are determined both by schools' initial situation and by long terms goals for 2030 that are identical for all schools within the same cycle. The incentive scheme involves a group bonus, where school staff receive a payment that is proportional to how much each school has improved compared to its target. In schools that have exactly attained their targets for example, staff receive a payment equivalent to 2.4 minimum wages while the upper limit for schools that have improved beyond their targets is 2.9 minimum wages[3]. The 10% best schools of the state of São Paulo also receive a reward regardless of their improvement[4]. Although teacher salaries vary depending on the grade and subject taught, and on teachers' qualifications, the bonus amounts on average to a little less than one monthly salary for a full-time middle school teacher, a relatively large payment compared to most other incentive programs analyzed in the literature.

São Paulo is not the only state in Brazil to implement initiatives aimed at rewarding good educational practices; however, in the majority of cases there is no direct link between teacher compensation and student performance. To my knowledge, only four other states so far (from a total of 27 states in Brazil), have implemented programs explicitly tying

[3]Only staff who worked at least 244 days during the year are eligible for the bonus.
[4]These rewards are lower than the regular bonus (1.5 minimum wages).

teacher bonus pay to student performance. In particular, the state of Pernambuco in northeast Brazil has adopted a very similar pay for performance program as the one in São Paulo in 2008. Other states that have recently implemented similar programs are Espírito Santo, Rio de Janeiro, and Amazonas[5].

2.3 Data

For the analysis in this paper I use data from *Prova Brasil*, a standardized test applied every two years to 5th and 9th grade students from all public schools[6], created in 2005. *Prova Brasil* is a low-stakes test designed as a tool to help teachers and policymakers assess the quality of education in Brazilian schools. In addition to taking an exam in math and language (Portuguese), students also provide socio-economic information through a survey. School principals and teachers also provide information about the school, its staff and working conditions through separate surveys.

The dataset used in this paper includes detailed information at the student level for 2007, 2009, 2011 and 2013, that is, one period of data before the implementation of the program, and three periods of data after the program was implemented. I also use 2007 School Census data to obtain information on total student enrollment (which allows the calculation of the percentage of test-takers in each school), and on the number of teachers by school. Columns 1-2 of Table 2.1 present descriptive statistics from the sample of 5th and 9th grade students before the implementation of the policy in 2007. Only students with available data on test scores are included, and who are enrolled in schools managed by the state or municipal government (students from other types of school represent less than 0.01% of the total). The share of test-takers relative to the total number of enrolled students is high, averaging around 90% for 5th grade students and 80% for 9th grade students. The data show that schools from the state of São Paulo are larger on average than schools from the rest of Brazil, and that their student population comes from a slightly more privileged socio-economic background, with a higher proportion of students whose parents finished high school.

In São Paulo as in the rest of Brazil, the majority of schools offering 5th grade education is managed by the municipal government, while the majority of schools offering 9th grade education is managed by the state government. However, during the period of analysis

[5]The bonus program in Espírito Santo, implemented in 2011, determines that teachers and school staff can receive up to one additional monthly salary per year based on student performance and other indicators such as absences. In Rio de Janeiro, where a similar program was implemented the same year, teachers can receive up to three monthly salaries. In Amazonas, a program implemented in 2008 determines that teachers from the best schools can receive up to two additional salaries, although its rules are different and a smaller proportion of schools is awarded bonuses.

[6]Only schools with a least 20 students enrolled in each grade are tested, and a very small number of private schools were tested during the period of analysis.

some schools switch from being run by the state government to being run by the municipal government, following a trend toward the decentralization of school management started in the 1990's. Overall, the share of municipal schools increases by around 10 percentage points between 2007 and 2013. As a result, some schools initially managed by the state government in São Paulo stop participating from the teacher pay for performance program in the course of this period. As most of these schools adopt new school identifiers, it is not possible to distinguish them from schools that drop out of the sample due to attrition. Given that the specific rules of the decentralization process might differ across states, I only include in the analysis schools that have not changed management in the period 2007-2013 to avoid selection issues[7]. Columns 3-4 of Table 2.1 show statistics for the final sample which only includes schools that have all 4 periods of data available. The new sample is about 70% the size of the original sample of schools offering 5th grade education and about 80% of the original sample of schools offering 9th grade education, and presents very similar observable characteristics as the original sample.

2.4 Empirical Strategy

In the absence of experimental data, the difference-in-differences (DD) method provides a way of estimating the impact of a program under the assumption that the outcomes of interest of the treated and control groups would have followed parallel trends over time in the absence of the treatment. If this identifying hypothesis is valid, the method allows the estimation of the average treatment effect on the treated (ATT). However, the estimator will be biased if there are time-varying unobserved factors that affect the outcome of both groups differently.

A typical concern in DD designs is the motivation behind the decision to participate in a program, as individuals who decide to enter a program may have specific unobserved characteristics that can affect their outcomes' trajectory. In the case of the São Paulo teacher incentive program, the participation decision has been taken at a centralized level, and concerns all state schools from the state of São Paulo regardless of their characteristics. It is still possible, however, that São Paulo state schools share distinct unobservable characteristics that would have led them to experience a different trend in performance compared to schools from other states in the absence of the program. Although the parallel trend hypothesis is impossible to verify, since only treated or control schools can be observed at a given time, I use a series of different counterfactuals in the estimations to test the robustness of the results.

First, I use students from São Paulo municipal schools as the comparison group, as they

[7] The results of the analysis are very similar, however, when using the whole sample of schools.

are not affected by the policy but are likely to be subject to the same state-specific time trends as students from São Paulo state schools. I then use students from state schools in other Brazilian states and regions to form several other comparison groups. The fact that there are both state and municipal schools within each state allows for a more robust analysis by using a triple differences framework (DDD), in order to remove state-specific time trends that could bias simple difference-in-differences estimations. Intuitively, the triple differences estimator is obtained by subtracting from the DD estimator differences in trends between São Paulo municipal schools and municipal schools from other states. If, however, there are trends that are specific to states and to the type of school, this method would not be effective in removing all bias.

The data available allow for a panel data analysis at the school level, as well as a student-level analysis using repeated cross-sections. I present the main results using cross sectional data at the student level with school fixed effects, as this allows for an analysis at a more disaggregated level and a larger number of observations[8]. School-level estimations provide very similar results and are used subsequently for further analysis.

Formally, the difference-in-differences equation can be written as follows:

$$Y_{ijt} = \alpha + \beta X_{ijt} + \sum_{\tau=2009}^{2013} \gamma_\tau \mathbb{1}(Year = \tau) + \sum_{\tau=2009}^{2013} \delta_\tau Treat_j * \mathbb{1}(Year = \tau) + \mu_j + \epsilon_{ijt} \quad (2.1)$$

Where Y_{ijt} is the outcome of student i in school j and year t, X_{ijt} is a vector of student-level covariates, $\mathbb{1}(Year = \tau)$ are a set of year dummies, $Treat_j$ is a dummy indicating treatment status (which equals 1 for the treated group, composed of schools managed by the state government of São Paulo), and μ_j are school fixed effects. The coefficients of interest are δ_τ, which capture the effect of the interaction between the treatment dummy and dummies for years in which the program was implemented. The fact that three periods of data are available after the implementation of the program allows me to use a flexible specification where treatment effects can vary over time. I also run an alternative specification where the treatment effect is constant over time. School fixed effects are intended to capture school characteristics that are fixed in time and could potentially influence the outcome, which includes any factors that are common to all schools in the state of São Paulo.

Similarly, the triple-differences estimator can be obtained through the following regression:

[8]Hanushek, Rivkin and Taylor (1996) discuss aggregation bias in analyses of student performance and suggests that aggregation can increase bias in some cases.

$$Y_{ijt} = \alpha + \beta X_{ijt} + \sum_{\tau=2009}^{2013} \gamma_\tau \mathbb{1}(Year = \tau) + \sum_{\tau=2009}^{2013} \delta_\tau SP_j * \mathbb{1}(Year = \tau)$$
$$+ \sum_{\tau=2009}^{2013} \theta_\tau SS_j * \mathbb{1}(Year = \tau) + \sum_{\tau=2009}^{2013} \phi_\tau SP_j * SS_j * \mathbb{1}(Year = \tau) + \mu_j + \epsilon_{ijt} \quad (2.2)$$

In addition to year dummies, the DDD equation includes interactions between year dummies and SP_j, a dummy indicating whether the school is located in the state of São Paulo, and interactions between year dummies and SS_j, a dummy indicating whether the school is managed by a state government. The three coefficients of interest are ϕ_{2009}, ϕ_{2011}, ϕ_{2013}, associated with the triple interaction, and as in the previous case I also run an alternative specification where the treatment effect is constant over time.

To assess whether student composition effects are driving the results (which might be the case for example if low-performing students are encouraged to dropout or change schools), I present estimates both with and without student-level covariates.

A potential concern with DD estimates is the fact that standard errors may be correlated within groups and over time, which could bias their estimation and understate the standard deviation of the estimator. This issue has been discussed by Bertrand, Duflo and Mullainathan (2004), who suggest that when the number of groups is large, one solution is to allow for the auto-correlation of standard errors. Following this idea, I cluster standard errors at the school level in all specifications.

2.5 Main Results

2.5.1 Student Performance

Tables 2.2 and 2.3 present the main results for 5th and 9th grade students respectively. In columns 1-2, I present simple DD estimates where the comparison group is composed by municipal schools in the state of São Paulo. Next, I present DDD estimates using a series of different counterfactuals from other Brazilian states. In columns 3-4 the comparison group is composed of all Brazilian state schools excluding São Paulo, and in columns 5-6 the comparison group is limited to state schools located in São Paulo adjacent states. Next, I only include in the analysis treated and control schools located in regions that are geographically close to each other, in an attempt to restrict the comparison to schools that are more similar in terms of unobservable characteristics.

In columns 7-8 only state schools located in "micro-regions"[9] that are at the boundary between São Paulo and its neighbor states are included in the analysis. Similarly, in columns 9-10 only state schools located in municipalities that are at the boundary between São Paulo and its neighbor states are included in the analysis. In all DDD regressions, I exclude states that have had pay for performance programs over the period of analysis as mentioned previously (Rio de Janeiro, Espírito Santo, Amazonas, and Pernambuco).

For a large proportion of students, basic socio-economic information is not available, and as a result when including covariates in the regression the sample size drops by more than half. In order to disentangle selection effects related to the availability of data from the effect of introducing covariates, I also run estimates without covariates using the restricted sample of students for which covariates are available. While I do not include these estimates in the tables for simplicity, they show that in cases where coefficients change with the introduction of covariates, the variation comes mainly from selection effects related to the availability of data, and that once we restrict the sample to students with basic socio-economic information, the coefficients remain practically unchanged when including covariates.

The estimated coefficients indicate that the program had overall positive and significant effects for 5th grade students. The size of the treatment effect varies according to the counterfactual chosen: overall gains in math range from 0.04 to 0.24 standard deviations, averaging 0.15 standard deviations, while gains in language are more modest, varying from close to zero to 0.14 standard deviations, and averaging 0.09 standard deviations. Interestingly, the estimated coefficients tend to increase as I progressively restrict the sample to geographically closer regions, suggesting unobserved factors might be leading to an underestimation of the effect size when using larger samples. The estimated effects of the program are more modest for 9th grade students, and the coefficients obtained are not statistically significant in several specifications. Improvements in math average 0.06 standard deviations, and overall effects in language average 0.03 standard deviations, although in the latter case coefficients are close to zero in most specifications. Overall, the estimated program impact is not constant across years: the coefficients obtained are generally smaller in 2011 than in 2009, but tend to increase again in 2013.

Tables 2.A1 and 2.A3 in the Appendix present simple DD estimates using the same counterfactuals used in previous triple differences estimations. These estimates show the coefficients obtained without taking into account the differential trends in student performance across states. Estimates for 5th grade students show positive and significant coefficients in most cases, although the coefficients are negative when using neighbor states as the comparison group. Meanwhile, simple DD estimates for 9th grade students

[9]Micro-regions are administrative divisions which include groups of municipalities based on socio-economic similarities. The state of São Paulo has 645 municipalities, grouped in 63 micro-regions.

show modest negative effects for both math and language test scores. Trends in student performance from these counterfactual groups are shown in Figures 2.A1-2.A5 in the Appendix.

Tables 2.A2 and 2.A4 in the Appendix show placebo DD estimates using the same counterfactuals, but assuming municipal schools in São Paulo are the treated group. Results show that over the period of analysis the performance of municipal schools from the state of São Paulo has evolved less favorably than the performance of municipal schools from other states. The magnitude of the estimated coefficients is similar for 5th and 9th grade students. This highlights the importance of using triple differences to account for state-specific trends that are unrelated to the program and may bias simple DD estimates. Final DDD estimates are obtained by subtracting placebo DD estimates from simple DD estimates, and provide more stable coefficients across counterfactuals.

2.5.2 Placebo Test

In this subsection, I use 2005 *Prova Brasil* data to check whether São Paulo state schools were already experiencing larger improvements in student performance relative to the counterfactuals groups considered prior to the implementation of the teacher incentive program. A limitation of this exercise, however, is that I only have access to data for 5th grade students in 2005. By using the same empirical strategy and sample of schools as in previous estimations, I estimate a placebo treatment effect assuming the program took place in 2007. The results, shown in Table 2.A5 in the Appendix, indicate that the performance of 5th grade students in São Paulo state schools deteriorated by around 0.1 to 0.3 standard deviations over this period depending on the counterfactual group used, with similar coefficients for math and language. This might indicate that previous results are a lower bound for the actual impact of the program, although it is not possible to rule out the possibility that this deterioration was related to a specific occurrence.

2.6 Heterogeneous Effects

2.6.1 Number of Teachers and Free-Riding Effects

The theoretical literature on group incentives suggests that free-riding behavior is more likely to occur as the group size increases. In large groups, each individual has less incentive to make efforts as his own impact on the overall outcome is lower, and given that he can expect to benefit from the effort of other workers. Goodman and Turner (2013) present evidence on this mechanism in the context of a group-based teacher incentive program in New York, and show that although the program was ineffective in general, schools with a smaller number of teachers experienced a modest increase in student achievement. Imberman and Lovenheim (2015) also provide evidence on the presence of free-riding among teachers participating in a group incentive program in Houston. They argue that the higher the share of students that a given teacher instructs, the stronger incentives become, as this increases teachers' impact on the overall outcome and reduces incentives to free ride. Accordingly, they show that student performance improved more among students whose teachers taught a larger share of students.

I take a similar approach as Goodman and Turner (2013), and look at whether the number of teachers teaching a specific grade affects the impact of the São Paulo teacher incentive program. Information on the number of teachers is obtained from 2007 School Census data. In 2007, the average number of 5th and 9th grade teachers in São Paulo state schools was 14 and 34 respectively. In order to assess how the number of teacher affects the impact of the program, I divide treated schools into four quartiles according to their number of teachers in 2007, and create dummies for three of the quartiles (d2, d3 and d4). I then interact each of these dummies with the treatment effect. This method has the advantage of giving more complete information than simply interacting the treatment variable with the number of teachers, while allowing other coefficients to be the same for all schools. One limitation of this approach is that there can be other factors related to the number of teachers which also affect student performance, although the inclusion of school fixed effects in the estimations already controls for the influence of school size.

Results are shown in Table 2.4. For simplicity, I only present DD estimates using municipal schools from the State of São Paulo as the comparison group, as estimates are very similar using other counterfactuals. The first line shows the overall program effects for reference, while the other coefficients are obtained from estimating equation (1) with the inclusion of the interacted variables mentioned above (the number of teachers is already captured by school fixed effects and is not included in the regressions). The estimates show that the number of teachers does not significantly impact the performance of 5th grade students. However, the performance of 9th grade students seems to be negatively affected by the number of teachers in a given school, although the estimated effects are modest. While 9th grade students from schools in the lowest quartile show no improvement in language test scores and a modest improvement of 0.03 standard deviations in math, the performance of those in the highest quartile deteriorates by 0.05-0.06 standard deviations (obtained by subtracting the highest quartile coefficient to the reference group coefficient).

2.6.2 Heterogeneity by Previous School Performance

In this section I assess whether the program's impact varies depending on schools' previous performance. The fact that the design of the program determines that rewards are proportional to schools' improvement (and not only given to schools that attain a discrete performance threshold), could be expected to lead to improvements in the entire distribution of schools, according to Lazear (2003). Nevertheless, it is possible that the program does not affect all schools equally if for example initially low-performing schools have more scope for improvement, or on the contrary lack the skills and capacity to respond to the program and improve performance.

To assess how treatment effects differ according to schools' initial performance, I take a similar approach as in the previous section and divide treated schools into four quartiles according to the distribution of test scores in 2007. Quartile dummies are then interacted with the treatment dummy. Table 2.5 reports the estimated coefficients. As in previous estimations, I only present results using municipal schools from the State of São Paulo as the counterfactual group.

The coefficients point to considerably different effects according to schools' initial performance. In the case of 5th grade students, while schools in the lowest quartile show an improvement of around 0.2 standard deviations in language scores, this effect decreases for schools in the following quartiles and drops to close to zero for schools in the highest quartile. A similar trend is observed for 5th grade math outcomes, although all types of schools experience positive gains in this case. Meanwhile, schools serving 9th grade students in the lowest quartile show gains of 0.07-0.08 standard deviations in both math and language, while schools in the highest quartile experience a deterioration in performance of around 0.1 standard deviations.

2.7 Conclusion

Despite being highly controversial, monetary incentives for teachers are an increasingly popular policy. However, the effects of this type of policy are not clear from a theoretical perspective and empirical evidence is limited so far. This paper contributes to the literature on teacher incentives and school accountability by studying the impact of a large teacher incentive program in the State of São Paulo, Brazil, which awarded bonuses at the group level for teachers and school staff conditional on improvements in student performance.

The results suggest the program had overall positive effects in performance, but that gains were more modest and less robust across specifications for 9th grade students than for 5th grade students. A potential explanation for this finding lies in the fact that 5th grade students generally have one main teacher, while 9th grade students have different teachers for different subjects taught, which might make coordination more difficult in the context of a group incentive program. Additionally, it may be easier to improve the performance of younger students, while older students may have learning gaps that are more difficult to close. The results also point to stronger achievement gains in math than in language for both grades, in line with findings from Muralidharan and Sundararaman (2011) and Lavy (2009), suggesting it might be more difficult to improve language skills in the short-term.

The fact that the magnitude of the estimated coefficients varies across specifications points to the difficulty of finding an appropriate counterfactual in difference-in difference analysis and the importance of dealing with potential confounding effects. In this study, I try to deal with state-specific trends that could potentially affect results through triple differences estimations, and assess the robustness of the results by using a series of different comparison groups.

The results also shows that the impact of the incentive program varies according to schools' initial characteristics. For 9th grade students, the number of teachers in a given school is associated with lower gains from the program, although the estimated effects are modest. More sizeable differences are found according to school's previous performance. Initially low-performing schools improved much more than the average, suggesting there may be considerable differences in the ability of schools to respond to this type of policy. Further research is needed to understand the factors behind the heterogeneity in gains among grades, subjects, and schools' initial characteristics; so that teacher incentive programs can be better targeted and more efficient in the future.

References

Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1):95–135.

Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H., and Wilson, D. (2009). Evaluating the impact of performance-related pay for teachers in England. *Labour Economics*, 16(3):251–261.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, 119(1)(February):249–75.

Besley, T. and Ghatak, M. (2005). Competition and Incentives with Motivated Agents. *American Economic Review*, 95(3):616–636.

Fryer, R. (2013). Teacher Incentives and Student Achievement: Evidence from New York City Public Schools. *Journal of Labor Economics*, 31(2).

Glazerman, S. and Seifullah, A. (2012). An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after Four Years. Final Report. *Mathematica Policy Research, Inc.*, pages 1–106.

Glewwe, P., Ilias, N., and Kremer, M. (2010). Teacher Incentives. *American Economic Journal: Applied Economics*, 2(3):205–227.

Goodman, S. F. and Turner, L. J. (2013). The Design of Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program. *Journal of Labor Economics*, 31(2):409–420.

Hanushek, E. a., Rivkin, S. G., and Taylor, L. L. (1996). Aggregation and the Estimated Effects of School Resources. *The Review of Economics and Statistics*, 78(4):611–627.

Holmstrom, B. (1982). Moral Hazard in Teams. *The Bell Journal of Economics*, 13(2):324–340.

Holmstrom, B. and Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization*, 7(Special Issue):24–52.

Imberman, S. A. and Lovenheim, M. F. (2015). Incentive Strenght and Teacher Productivity: Evidence from a Group-based Teacher Incentive Pay System. *The Review of Economic Studies*, 97(2):364–386.

Lavy, V. (2002). Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement. *Journal of Political Economy*, 110(6):1286–1317.

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, 99(5):1979–2011.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review*, 10:179–214.

Martins, P. S. (2009). Individual Teacher Incentives, Student Achievement, and Grade Inflation. *Applied Economics*, (4051):43.

Muralidharan, K. and Sundararaman, V. (2011). Teacher Performance Pay : Experimental Evidence from India. *Journal of Political Economy*, 119(1):39–77.

Neal, D. (2011). The Desgin of Performance Pay in Education. *Handbook of Economics of Education*, Vol. 4.

Oshiro, C. H., Scorzafave, L. G., and Dorigan, T. A. (2015). Impacto Sobre o Desempenho Escolar do Pagamento de Bônus aos Docentes do Ensino Fundamental do Estado de São Paulo. *Revista Brasileira de Economia*, 69(2):213–249.

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

Rockoff, J. E. . (2004). The Impact of Individual Teachers on Student Achievement : Evidence from Panel Data. *American Economic Review*, 94(2):247–252.

Springer, M. G., Dale, B., Hamilton, L., Le, V.-N., Lockwood, J., McCaffrey, D., Pepper, M., and Stecher, B. M. (2010). Teacher Pay for Performance Exprimental Experimental Evidence from the Project on Incentives Teaching. *National Center on Performance Incentives, Project on Incentives in Teaching.*

Tables and Figures

Table 2.1: Descriptive statistics (year 2007)

	Brazil (all schools) (1)	São Paulo (all schools) (2)	Brazil (schools w/ 4 years of data) (3)	São Paulo (schools w/ 4 years of data) (4)
5th grade				
No. of test-takers	2,293,687	580,675	1,809,660	441,555
No. of schools	37,104	5,589	25,972	4,120
Avg. test-takers per school	62	104	70	107
% test-takers relative to total	90	92	91	92
% state schools	33	40	30	41
% municipal schools	67	60	70	59
Avg. age	10.8	10.4	10.8	10.4
% black	12	9	12	9
% father w/ high school	34	40	34	40
% mother w/ high school	31	36	32	36
Avg. Test score - language (0-100)	50	52	50	52
Avg. Test score - math (0-100)	51	53	51	54
9th grade				
No. of test-takers	1,785,895	494,487	1,546,368	467,442
No. of schools	27,163	4,666	21,273	4,308
Avg. test-takers per school	66	106	73	109
% test-takers relative to total	82	84	81	84
% state schools	69	80	69	80
% municipal schools	31	20	31	20
Avg. age	15.4	15	15.3	15
% black	11	10	11	10
% father w/ high school	30	34	31	35
% mother w/ high school	31	32	32	32
Avg. Test score - language (0-100)	57	58	57	58
Avg. Test score - math (0-100)	57	57	57	57

Note: 5th grade students are graded in a scale of 0-350 for language (Portuguese) and 0-375 for math, while 9th grade students are graded in a scale of 0-400 for Portuguese and 0-425 for math. For ease of comparability, grades are rescaled in a range of 0-100.

Table 2.2: Difference-in-differences and triple differences estimates - 5th grade

Comparison group	São Paulo municipal schools (DD)		Rest of Brazil (DDD)		Neighbor states (DDD)		Neighbor micro-regions (DDD)		Neighbor municipalities (DDD)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Language (Portuguese)										
Overall effect	0.086***	0.111***	0.093***	0.065***	0.108***	-0.007	0.125***	0.082**	0.140***	0.120**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.04)	(0.04)	(0.06)
2009	0.098***	0.119***	0.110***	0.078***	0.104***	-0.018	0.124***	0.079*	0.131***	0.106
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.05)	(0.07)
2011	0.051***	0.073***	0.068***	0.033***	0.081***	-0.031*	0.077**	0.032	0.122**	0.096
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.05)	(0.05)	(0.07)
2013	0.117***	0.153***	0.102***	0.087***	0.146***	0.038**	0.184***	0.139***	0.172***	0.166**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.05)	(0.06)	(0.07)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	1,567,525	659,445	5,843,338	2,226,555	3,017,941	1,202,437	478,137	198,413	188,325	76,653
Math										
Overall effect	0.120***	0.152***	0.116***	0.095***	0.135***	0.042***	0.210***	0.181***	0.242***	0.244***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.05)	(0.06)
2009	0.129***	0.156***	0.117***	0.091***	0.104***	0.005	0.187***	0.185***	0.238***	0.263***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.05)	(0.06)	(0.07)
2011	0.086***	0.115***	0.101***	0.072***	0.125***	0.037**	0.175***	0.139***	0.246***	0.251***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.05)	(0.06)	(0.07)
2013	0.153***	0.203***	0.134***	0.134***	0.188***	0.106***	0.278***	0.222***	0.242***	0.208***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.05)	(0.06)	(0.06)	(0.08)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	1,567,287	659,425	5,842,494	2,226,402	3,017,504	1,202,372	478,069	198,399	188,307	76,650

Note: Outcomes are standardized test scores. All regressions use school and year fixed effects. Student level controls include: gender, race (a dummy=1 for black students), mother and father education (a dummy=1 if the mother/father have completed high school). Standard errors clustered at the school level in parentheses.

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.3: Difference-in-differences and triple differences estimates - 9th grade

Comparison group	São Paulo municipal schools (DD)		Rest of Brazil (DDD)		Neighbor states (DDD)		Neighbor micro-regions (DDD)		Neighbor municipalities (DDD)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Language (Portuguese)										
Overall effect	-0.027***	-0.023***	-0.002	-0.006	0.000	-0.006	0.119***	0.117***	0.078	0.066
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.06)	(0.06)
2009	-0.015	-0.011	-0.010	-0.012	-0.043***	-0.041**	0.107**	0.112**	0.085	0.074
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(0.05)	(0.07)	(0.06)
2011	-0.053***	-0.054***	-0.031***	-0.044***	0.032*	0.013	0.073*	0.043	0.044	-0.001
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.05)	(0.07)	(0.07)
2013	-0.012	-0.006	0.037***	0.035***	0.013	0.017	0.181***	0.193***	0.108	0.123
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(0.06)	(0.08)	(0.08)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	1,796,341	1,190,513	5,398,627	3,387,310	3,103,450	2,032,720	472,196	320,711	182,931	123,970
Math										
Overall effect	-0.012	-0.008	0.023**	0.018*	0.039***	0.031**	0.150***	0.136***	0.105*	0.080
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.04)	(0.06)	(0.06)
2009	0.006	0.006	0.022*	0.017	0.010	0.007	0.117**	0.110**	0.104*	0.081
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(0.05)	(0.06)	(0.06)
2011	-0.051***	-0.053***	-0.017	-0.029**	0.048***	0.032*	0.087*	0.053	0.044	-0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(0.05)	(0.07)	(0.07)
2013	0.011	0.018	0.064***	0.064***	0.060***	0.059***	0.246***	0.244***	0.171**	0.164**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(0.06)	(0.08)	(0.08)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	1,796,295	1,190,496	5,398,269	3,387,163	3,103,324	2,032,677	472,192	320,711	182,928	123,976

Note: Outcomes are standardized test scores. All regressions use school and year fixed effects. Student level controls include: gender, race (a dummy=1 for black students), mother and father education (a dummy=1 if the mother/father have completed high school). Standard errors clustered at the school level in parentheses.

$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Table 2.4: Heterogeneous effects by number of teachers

Comparison group	São Paulo municipal schools (DD)			
	5th grade		9th grade	
	(1)	(2)	(3)	(4)
	Language (Portuguese)			
Overall effect	0.093***	0.098***	-0.021**	-0.019**
	(0.01)	(0.01)	(0.01)	(0.01)
Treatment (ref. group=1)	0.091***	0.098***	0.011	0.009
	(0.01)	(0.01)	(0.01)	(0.01)
Treatment*d2	-0.008	-0.015	-0.021*	-0.015
	(0.02)	(0.02)	(0.01)	(0.01)
Treatment*d3	0.008	0.015	-0.045***	-0.042***
	(0.02)	(0.02)	(0.01)	(0.01)
Treatment*d4	0.009	0.003	-0.056***	-0.053***
	(0.01)	(0.01)	(0.01)	(0.01)
Controls	No	Yes	No	Yes
Obs.	16,477	16,473	17,227	17,226
	Math			
Overall effect	0.132***	0.137***	-0.007	-0.004
	(0.01)	(0.01)	(0.01)	(0.01)
Treatment (ref. group=1)	0.125***	0.132***	0.025**	0.026**
	(0.01)	(0.01)	(0.01)	(0.01)
Treatment*d2	0.033	0.026	-0.021*	-0.017
	(0.02)	(0.02)	(0.01)	(0.01)
Treatment*d3	-0.021	-0.013	-0.050***	-0.045***
	(0.02)	(0.02)	(0.01)	(0.01)
Treatment*d4	-0.009	-0.014	-0.057***	-0.055***
	(0.02)	(0.02)	(0.01)	(0.01)
Controls	No	Yes	No	Yes
Obs.	16,477	16,473	17,227	17,226

Note: Outcomes are standardized test scores. All regressions use school and year fixed effects. School level controls include: % of girls, % of black students, % of students whose mother completed high school, % of students whose father completed high school. Only schools with at least 10 test-takers are included in the regressions. Standard errors clustered at the school level in parentheses. The number of observations range from 16,319 to 17,227 for 9th grade estimations, and from 15,845 to 16,477 for 5th grade estimations.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.5: Heterogeneous effects by initial school performance

Comparison group	São Paulo municipal schools (DD)			
	5th grade		9th grade	
	(1)	(2)	(3)	(4)
	Language (Portuguese)			
Overall effect	0.093***	0.098***	-0.021**	-0.019**
	(0.01)	(0.01)	(0.01)	(0.01)
Treatment (ref. group=1)	0.213***	0.215***	0.088***	0.072***
	(0.01)	(0.01)	(0.01)	(0.01)
Treatment*d2	-0.116***	-0.113***	-0.095***	-0.080***
	(0.02)	(0.02)	(0.01)	(0.01)
Treatment*d3	-0.136***	-0.136***	-0.131***	-0.108***
	(0.02)	(0.01)	(0.01)	(0.01)
Treatment*d4	-0.226***	-0.220***	-0.209***	-0.180***
	(0.02)	(0.02)	(0.01)	(0.01)
Controls	No	Yes	No	Yes
Obs.	16,477	16,473	17,227	17,226
	Math			
Overall effect	0.132***	0.137***	-0.007	-0.004
	(0.01)	(0.01)	(0.01)	(0.01)
Treatment (ref. group=1)	0.212***	0.218***	0.083***	0.076***
	(0.01)	(0.01)	(0.01)	(0.01)
Treatment*d2	-0.057***	-0.058***	-0.065***	-0.057***
	(0.02)	(0.02)	(0.01)	(0.01)
Treatment*d3	-0.094***	-0.099***	-0.117***	-0.102***
	(0.02)	(0.02)	(0.01)	(0.01)
Treatment*d4	-0.169***	-0.168***	-0.180***	-0.162***
	(0.02)	(0.02)	(0.01)	(0.01)
Controls	No	Yes	No	Yes
Obs.	16,477	16,473	17,227	17,226

Note: Outcomes are standardized test scores. All regressions use school and year fixed effects. School level controls include: % of girls, % of black students, % of students whose mother completed high school, % of students whose father completed high school. Only schools with at least 10 test-takers are included in the regressions. Standard errors clustered at the school level in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Appendix to Chapter 2

Table 2.A1: Simple difference-in-differences estimates - 5th grade

Comparison group	Rest of Brazil		Neighbor states		Neighbor micro-regions		Neighbor municipalities	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Language (Portuguese)				
Overall effect	0.029***	0.042***	-0.074***	-0.055***	0.009	0.029	0.030	0.063
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.05)
2009	0.064***	0.078***	-0.055***	-0.050***	0.027	0.044	0.025	0.053
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.04)	(0.04)	(0.05)
2011	-0.035***	-0.018*	-0.116***	-0.085***	-0.053	-0.032	-0.018	0.014
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.04)	(0.04)	(0.05)
2013	0.064***	0.063***	-0.042***	-0.026*	0.059*	0.075*	0.094**	0.130**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.04)	(0.05)	(0.06)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	1,788,321	713,597	1,027,571	422,536	84,532	36,513	51,596	22,057
				Math				
Overall effect	0.104***	0.128***	-0.023**	0.006	0.105***	0.150***	0.143***	0.202***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.04)	(0.04)	(0.05)
2009	0.137***	0.157***	-0.030***	-0.016	0.105***	0.160***	0.132***	0.207***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)	(0.06)
2011	0.048***	0.076***	-0.054***	-0.02	0.057	0.104**	0.115***	0.180***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)	(0.06)
2013	0.131***	0.152***	0.028**	0.066***	0.164***	0.185***	0.192***	0.217***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.05)	(0.05)	(0.06)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	1,788,321	713,597	1,027,439	422,521	84,509	36,509	51,582	22,058

Note: Outcomes are standardized test scores. Only state schools are included in the regressions. All regressions use school and year fixed effects. Student level controls include: gender, race (a dummy=1 for black students) mother education, father education (a dummy=1 if the mother/father have completed high school). Standard errors clustered at the school level in parentheses.

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.A2: Placebo difference-in-differences estimates - 5th grade

Comparison group	Rest of Brazil		Neighbor states		Neighbor micro-regions		Neighbor municipalities	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Language (Portuguese)				
Overall effect	-0.063***	-0.022***	-0.180***	-0.048***	-0.116***	-0.051**	-0.109***	-0.055
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
2009	-0.046***	0.001	-0.159***	-0.032***	-0.097***	-0.035	-0.106***	-0.052
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.04)
2011	-0.104***	-0.051***	-0.197***	-0.054***	-0.130***	-0.062**	-0.140***	-0.081**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)
2013	-0.038***	-0.023**	-0.188***	-0.063***	-0.124***	-0.063**	-0.078**	-0.034
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	4,055,017	1,512,958	1,990,370	779,901	393,605	161,900	136,729	54,596
				Math				
Overall effect	-0.011	0.033***	-0.157***	-0.035***	-0.104***	-0.03	-0.085***	-0.027
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.04)
2009	0.020***	0.066***	-0.134***	-0.021*	-0.082***	-0.024	-0.084**	-0.034
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.04)	(0.04)
2011	-0.053***	0.006	-0.179***	-0.048***	-0.119***	-0.033	-0.119***	-0.057
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)
2013	-0.003	0.019*	-0.160***	-0.039***	-0.114***	-0.035	-0.05	0.012
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.05)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	4,054,419	1,512,850	1,990,065	779,851	393,560	161,890	136,725	54,592

Note: Outcomes are standardized test scores. Only municipal schools are included in the regressions. All regressions use school and year fixed effects. Student level controls include: gender, race (a dummy=1 for black students) mother education, father education (a dummy=1 if the mother/father have completed high school). Standard errors clustered at the school level in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.A3: Simple difference-in-differences estimates - 9th grade

Comparison group	Rest of Brazil		Neighbor states		Neighbor micro-regions		Neighbor municipalities	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Language (Portuguese)				
Overall effect	-0.039***	-0.038***	-0.068***	-0.059***	-0.018	-0.007	-0.027	-0.028
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
2009	-0.057***	-0.047***	-0.087***	-0.067***	-0.013	0.007	0.001	0.008
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
2011	-0.031***	-0.031***	-0.067***	-0.064***	-0.032**	-0.034**	-0.052**	-0.068**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
2013	-0.030***	-0.034***	-0.049***	-0.044***	-0.008	0.003	-0.03	-0.033
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	3,795,289	2,423,248	2,504,379	1,647,043	379,193	256,892	144,396	97,773
				Math				
Overall effect	-0.033***	-0.026***	-0.049***	-0.036***	0.004	0.016	0.013	0.01
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
2009	-0.039***	-0.034***	-0.054***	-0.040***	0.001	0.02	0.019	0.022
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
2011	-0.071***	-0.065***	-0.092***	-0.082***	-0.047***	-0.048***	-0.054**	-0.071**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
2013	0.011*	0.018***	0	0.013	0.064***	0.074***	0.077***	0.075**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	3,795,069	2,423,157	2,504,283	1,647,006	379,195	256,894	144,395	97,778

Note: Outcomes are standardized test scores. Only state schools are included in the regressions. All regressions use school and year fixed effects. Student level controls include: gender, race (a dummy=1 for black students) mother education, father education (a dummy=1 if the mother/father have completed high school). Standard errors clustered at the school level in parentheses.

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.A4: Placebo difference-in-differences estimates - 9th grade

Comparison group	Rest of Brazil		Neighbor states		Neighbor micro-regions		Neighbor municipalities	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Language (Portuguese)				
Overall effect	-0.037***	-0.031***	-0.069***	-0.053***	-0.137***	-0.123***	-0.106*	-0.093
	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.06)	(0.06)
2009	-0.046***	-0.035***	-0.044***	-0.026*	-0.120***	-0.104**	-0.084	-0.064
	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.04)	(0.06)	(0.06)
2011	0.000	0.013	-0.099***	-0.077***	-0.104**	-0.077*	-0.096	-0.065
	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.04)	(0.06)	(0.07)
2013	-0.067***	-0.068***	-0.063***	-0.060***	-0.188***	-0.189***	-0.138*	-0.153*
	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(0.05)	(0.08)	(0.08)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	1,603,338	964,062	599,071	385,677	93,003	63,819	38,535	26,197
				Math				
Overall effect	-0.056***	-0.044***	-0.088***	-0.066***	-0.145***	-0.118***	-0.092*	-0.069
	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.05)	(0.05)
2009	-0.061***	-0.051***	-0.064***	-0.046***	-0.116***	-0.088**	-0.085	-0.059
	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.04)	(0.06)	(0.05)
2011	-0.054***	-0.035***	-0.140***	-0.113***	-0.135***	-0.100**	-0.098	-0.064
	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.05)	(0.06)	(0.07)
2013	-0.054***	-0.045***	-0.060***	-0.045**	-0.183***	-0.168***	-0.094	-0.086
	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(0.05)	(0.07)	(0.08)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	1,603,200	964,006	599,041	385,671	92,997	63,817	38,533	26,198

Note: Outcomes are standardized test scores. Only municipal schools are included in the regressions. All regressions use school and year fixed effects. Student level controls include: gender, race (a dummy=1 for black students) mother education, father education (a dummy=1 if the mother/father have completed high school). Standard errors clustered at the school level in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.A5: Placebo estimates 2005-2007 - 5th grade

Comparison group	São Paulo municipal schools (DD)		Rest of Brazil (DDD)		Neighbor states (DDD)		Neighbor micro-regions (DDD)		Neighbor municipalities (DDD)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Language (Portuguese)										
Placebo effect	-0.201***	-0.204***	-0.230***	-0.206***	-0.336***	-0.272	-0.139***	-0.157**	-0.155***	-0.174***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.04)	(0.05)	(0.05)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	613,959	308,177	2,607,918	1,167,448	1,328,752	598,306	201,215	95,724	82,556	37,736
Math										
Placebo effect	-0.176***	-0.176***	-0.210***	-0.194***	-0.295***	-0.253***	-0.114**	-0.158***	-0.141**	-0.174**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.05)	(0.06)	(0.07)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	613,827	308,161	2,607,593	1,167,381	1,328,538	598,274	201,181	95,715	82,539	37,734

Note: outcomes are standardized test scores. All regressions use school and year fixed effects. Student level controls include: gender, race (a dummy=1 for black students), mother and father education (a dummy=1 if the mother/father have completed high school). Standard errors clustered at the school level in parentheses.

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Figure 2.A1: Trends in student performance - state and municipal schools in São Paulo

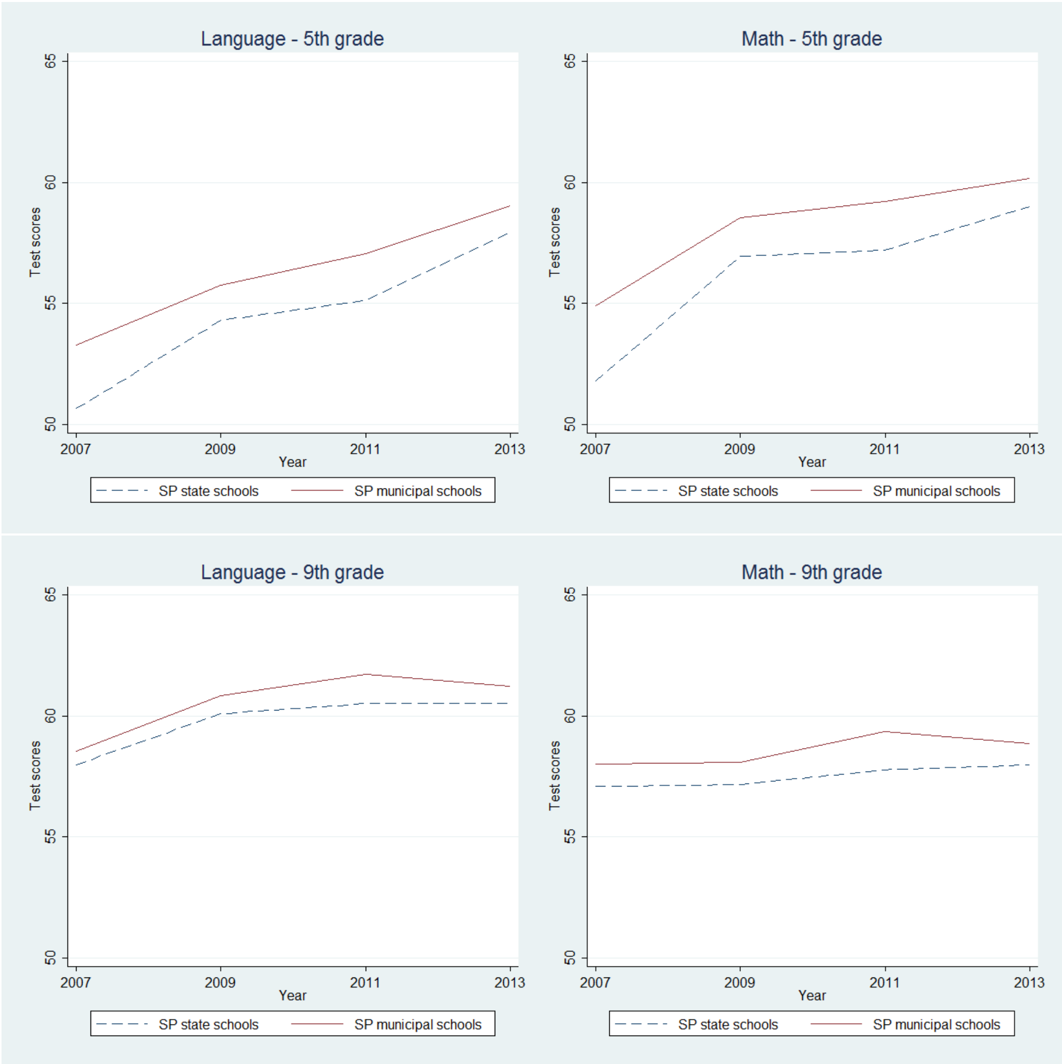


Figure 2.A2: Trends in student performance - state schools in São Paulo and the rest of Brazil

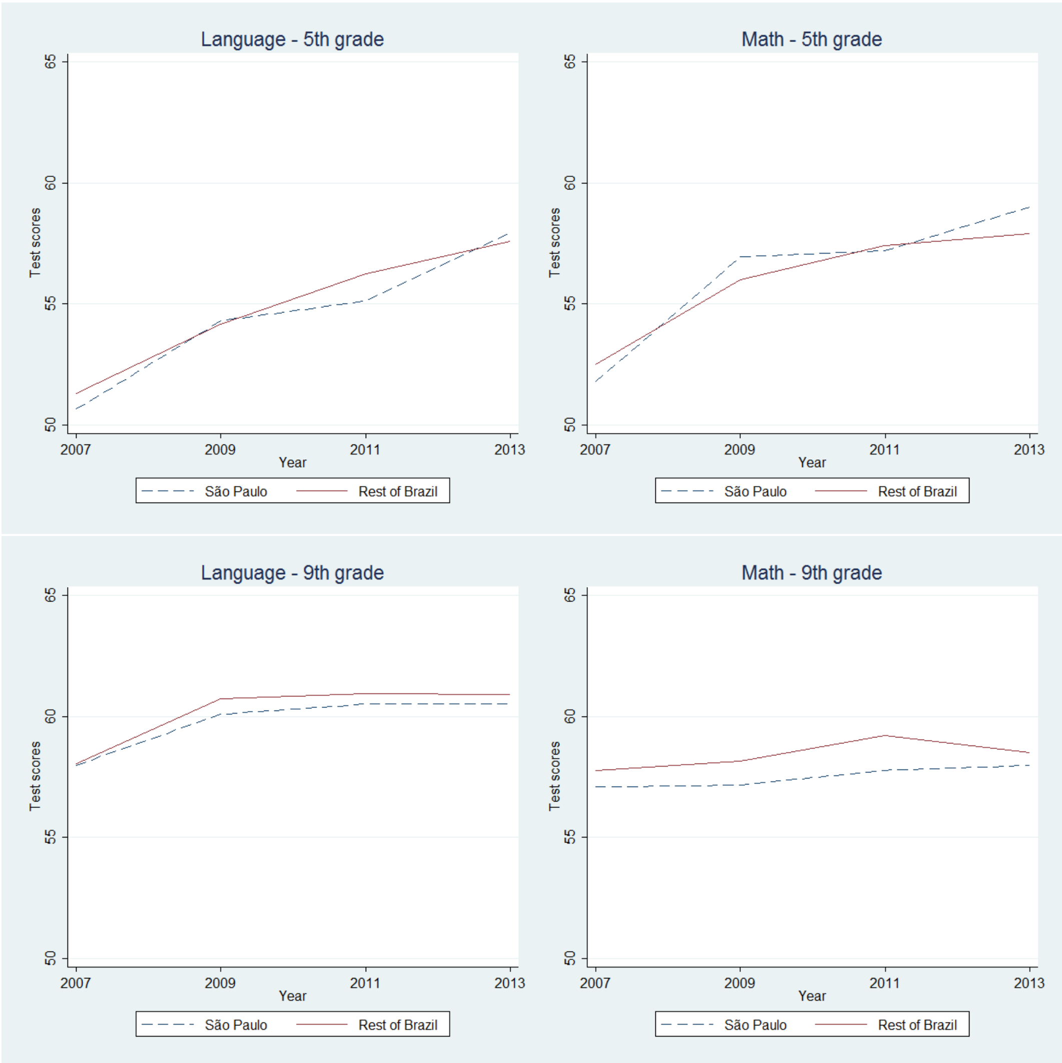


Figure 2.A3: Trends in student performance - state schools in São Paulo and neighbor states

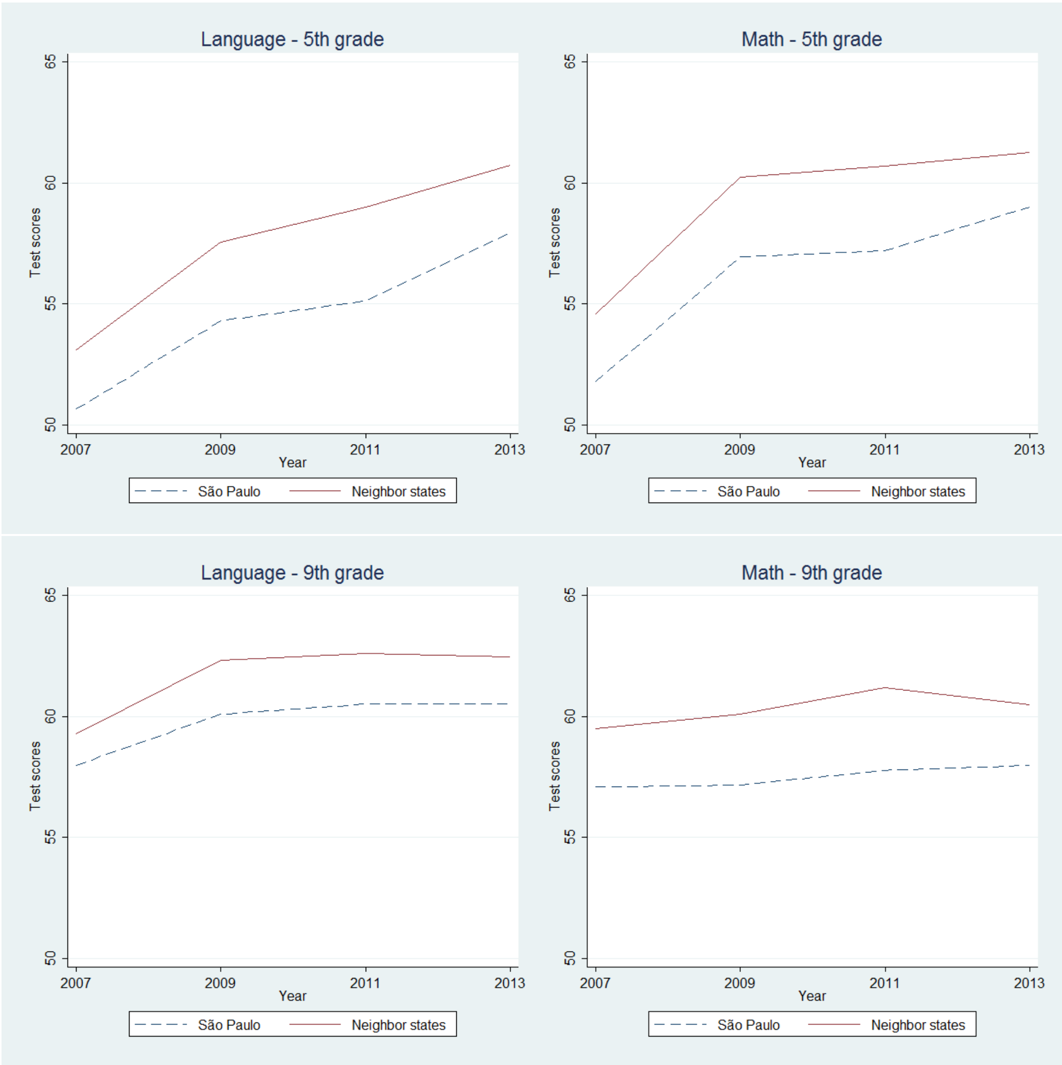


Figure 2.A4: Trends in student performance - state schools in São Paulo and neighbor states, adjacent micro-regions only

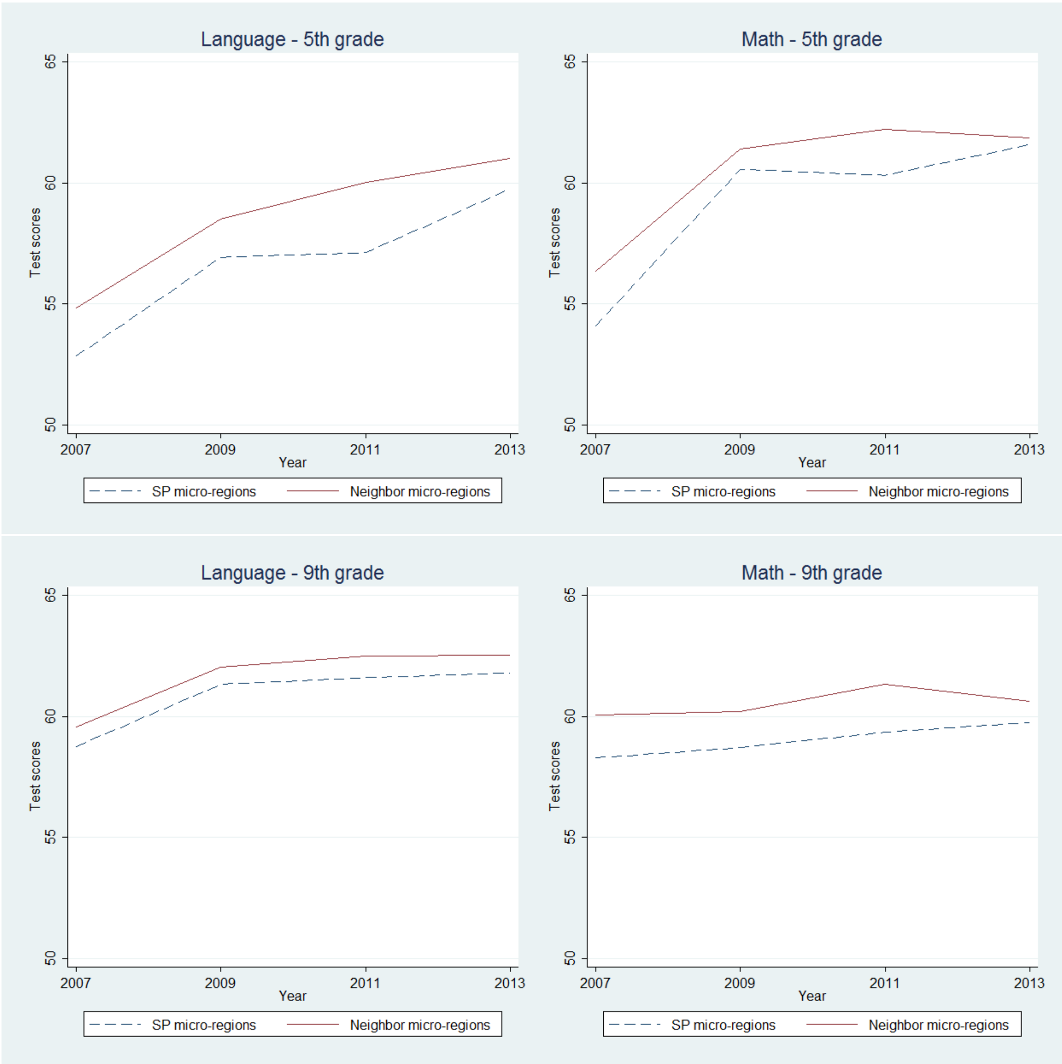
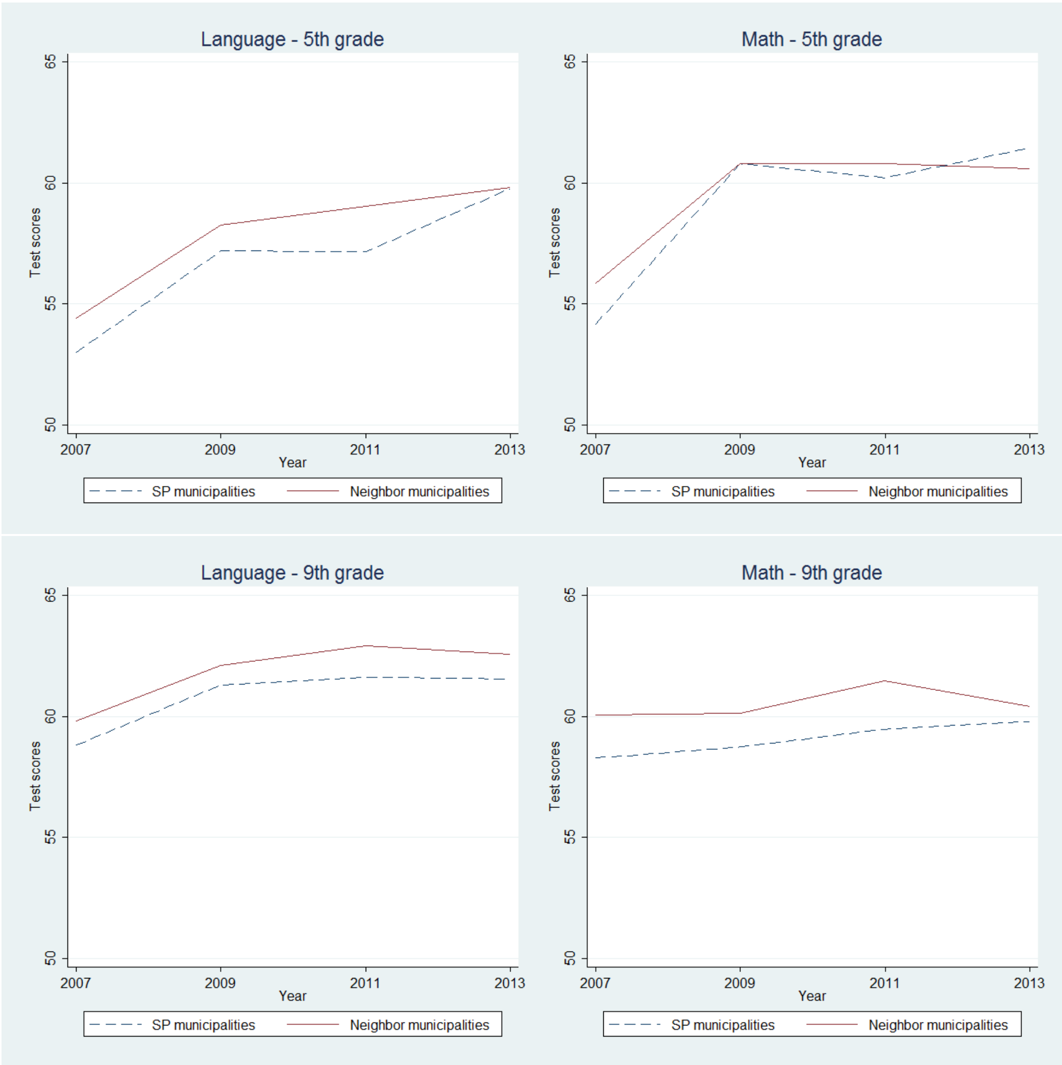


Figure 2.A5: Trends in student performance - state schools in São Paulo and neighbor states, adjacent municipalities only



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

3.1 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, which establish financial or political sanctions or rewards according to schools' performance. 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 school in a natural experiment in North Carolina. Similarly, a study in Canada by Friesen, Javdani, Smith and Woodcock (2012) shows 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 United States, 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].

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

[2] Although school test scores 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).

However, other studies show that school performance is not always the main determinant of choice and that preferences regarding schools are heterogeneous across socio-economic groups. Hastings, Kane and Staiger (2009) estimate a demand model for schools and show that high-income families place a higher weight on school quality, while social preferences play a larger role for families from minority or disadvantaged groups. Along the same lines, Elacqua, Schneider and Buckley (2007) show that parents value the social composition of the student body when choosing schools in Chile, and Gallego and Hernando (2010) 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, Das and Reis (2013) show 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 (Exame Nacional do Ensimo Médio)*,[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 provides a setting for a sharp regression discontinuity design. In order to allow for different effects according to schools' performance, I split the sample between high- 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. In a related study in

[3]In Brazil an indicator of school performance at the national level for basic education (Ideb - Índice de Desenvolvimento da Educação Básica) 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.

Brazil using *Enem* data, Camargo, Camelo, Firpo and Ponczek (2014) find that publishing test scores at the school level results in a subsequent improvement of school quality for private schools scoring below the median. Similarly, Koning and van der Wiel (2010) find that schools that received a low score and had their scores published by a newspaper in the Netherlands subsequently improved their grades. In Chile, Neilson (2013) tries to disentangle both effects in the context of a school voucher policy. Here I focus exclusively on the demand side, and only look at short term reallocation effects without taking into account possible changes in school quality.

I find that the coefficients associated with the discontinuity have the expected sign in almost all specifications: the effect of having grades disclosed on enrollment is positive for the best-performing schools, and negative for the worst-performing schools. However, these coefficients are not statistically significant for either private or public schools. In baseline estimations, the coefficients point to changes of up to 6% of enrolled students in either direction, 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 defining the sample of 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, Das and Khwaja (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 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

[5] Most papers study the effects of vouchers or "hard" accountability policies.

school is not well known.

The remainder of this paper is organized as follows. Section 1.2 provides background on the education system in Brazil and the *Enem* exam. Section 1.3 presents the data used and some descriptive statistics. Section 1.4 details the empirical strategy, and estimation results are reported in Section 1.5. A few robustness checks are reported in Section 1.6, and Section 1.7 presents some concluding remarks.

3.2 Background: the Education System in Brazil and the *Enem* Exam

3.2.1 The Education System in Brazil

The basic education system in Brazil is divided by cycles. After preschool, the second cycle of basic education lasts nine years and is attended by students aged 6 to 14 approximately (primary and middle school), while the third cycle (high school) lasts three years and is attended by students aged 15 to 17 approximately. Both private schools and free public schools coexist. Public schools account for over 85% of enrollments, as shown in Table 3.1, 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 regarding school choice. But even in states where admissions are centralized such as the State of São Paulo (the richest and most populous in Brazil), where children are encouraged to attend a school near their home, students can generally

[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.

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.

3.2.2 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, the first year the exam was implemented, 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* is used as one of the criteria for government scholarship attributions for disadvantaged college students since 2005 (through the *Prouni*[8] program), and many public 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, only students enrolled in the last year of high school are taken into account in the calculation of school grades.

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:

[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.

[8] *Programa Universidade Para Todos*

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[9]. *Enem* scores at the school level started being publicly released from 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 3.1 shows a few examples of published rankings, and Figure 3.2 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 3.2.

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 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.

3.3 Data

3.3.1 *Enem* Data

This paper uses *Enem* microdata for the years 2005 (the first session for which grades were publicly released at the school level) to 2008. I do not use data from 2009 onwards, as the exam format and publication rules have changed since then, as described previously. It is not clear whether students and their families continued to interpret test scores the same way after these changes 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

[9]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 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.

means the same regression discontinuity design cannot be applied.

Enem data provide information on each student that has taken the exam including test scores, 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. Only students enrolled in the final year of high school, and who were 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.

3.3.2 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, and to 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 taking place at t, 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[10]. Over 97% of schools from the original sample could be matched to school census 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 these observations as missing, which represent approximately 10% of the sample when pooling together all years of data[11].

[10] 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.

[11] Although these schools have slightly lower *Enem* averages than the rest of the sample, the grades do not differ significantly between schools from each side of the discontinuity in the window of data considered in the analysis.

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 point.

3.3.3 Descriptive Statistics

Table 3.3 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 more likely to have missing data than bigger schools[12]. 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 between both sides, as will be shown in the next section.

The analysis in this paper focuses on a specific subset of schools with a small number of test takers. Table 3.4 provides a view of how these schools differ from others (statistics shown are calculated using pooled data from different years). Schools with a smaller number of test takers are more often private, and are generally smaller. The share of

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

students who take the exam is also lower in these schools (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 who takes the exam in these schools, although their average grades are not very different from grades of larger schools.

3.4 Empirical Strategy

3.4.1 Methodology

The question this paper wants to address is whether schools' performance in the *Enem* exam has an effect on students' school choices, 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 stating that only schools with at least 10 test takers would have their results published. In order to allow effects to differ according to the type of school, I split the sample in order to create a group of high-performing schools and a group of low-performing schools. As the criterion used for splitting the sample is arbitrary, I use alternative 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 attention to data close enough to the discontinuity, we are in a case similar to that of 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 (2008), 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 restrict the analysis to a symmetrical window of data around the discontinuity of 5-15 *Enem* takers, and run OLS regressions separately on both sides of the discontinuity, which is the equivalent of estimating the following equation:

$$log(Y_{it+2}) = \beta_0 + \phi_1(X_{it} - c) + \beta_1 D_{it} + D_{it}\phi_2(X_{it} - c) + \epsilon_{it} \quad (3.1)$$

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 which equals 1 if school i has its grades published following the *Enem* session taking place 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 in every period, then school fixed effects would capture the effect of the publication of grades, and these schools would effectively be excluded from the analysis.

3.4.2 Internal Validity

The regression discontinuity design might be invalidated if schools are able to manipulate the number of students who take the exam (for example encouraging or discouraging students to take the *Enem*), and therefore influence whether or not their test scores 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 present the day of the exam that is taken into account by the publication rule, rather than 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 3.3, 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 3.4, does not show jumps around the threshold. This result holds for different choices of bandwidth and polynomial degrees, and also when looking separately at private and public schools (not shown here).

Another way of assessing whether there is manipulation by schools is to look for jumps in covariates around the discontinuity. If the only difference between treated and control schools 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 (3.1) where different covariates at the school level in year t are used as explanatory variables, in order to see whether the dummy coefficient β_1 is statistically significant.

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 3.5 generally suggest there are no jumps in covariates, although the dummy coefficient is significant for enrollment pre-treatment data when adding a quadratic term. Further investigation shows this result is not robust to minor specification changes such as adding a cubic term, or when splitting 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 year t in all specifications to deal with any possible confounding factors. I also run the same regressions separately for private and public schools (shown in Tables 3.A1 and 3.A2 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 3.6 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.

3.5 Results

3.5.1 Main Results

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 in particular as private schools have much more flexibility in their admissions procedures, I run separate regressions for each type of school. I estimate equation (3.1), where I also control for enrollment in year t 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 year t is particularly important since jumps in pre-enrollment data were found 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 create a group of 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 columns 1-2 of Table 3.7. The publication of school grades does not lead to a significant change in the number of enrolled students in either type of schools. However, the sign of the dummy coefficient goes in the expected direction in most cases, with high performing schools showing an increase in enrollments and low performing schools showing a decrease in enrollments 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 highest gain, of 11%-12% additional students, although not statistically significant.

It is possible 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 3.5, 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 significant effects for private schools.

To allow for this possibility, I create separate rankings for each type of school. Columns 3-4 and columns 5-6 of Table 3.7 present results where the samples include 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.

The fact that the sample size is relatively small for some subgroups means I may have low statistical power to detect any effects for these groups. 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 cutoffs and plot the coefficients obtained in a graph (Figure 3.6). Results do not suggest a clear pattern and in most cases coefficients fluctuate around zero, although high-performing public schools have consistently positive coefficients.

3.5.2 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 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 could be masking significant disparities, with reallocation effects only taking place in municipalities where schools face a 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 (3.2):

$$\begin{aligned} log(Y_{it+2}) = \beta_0 + \phi_1(X_{it} - c) + \beta_1 D_{it} + D_{it}\phi_2(X_{it} - c) \\ + \beta_2 Comp_{it} + \beta_3 D_{it} Comp_{it} + \epsilon_{it} \end{aligned} \tag{3.2}$$

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 1-4 of Table 3.8. 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[13].

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 using 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 columns 5-6 of Table 3.8 and do not alter previous conclusions.

3.5.3 Schools' Socio-Economic Environment

I next consider the socio-economic environment of schools at the level of municipalities. As mentioned in the Introduction, 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* [14]. I use data from the year 2000 for all indicators considered.

[13] Results are very similar when including polinomials for the interacted terms and competition measures.

[14] 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/

As in previous estimations, baseline regressions are run with a specification similar to equation (3.2), where the treatment dummy is interacted with socioeconomic variables at the municipality level, designed by SE_{it}. Estimation results are presented in Table 3.9, 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.

3.6 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 Tables 3.A3 and 3.A4 in the Appendix and are very similar to those obtained previously. Second, I run the same regressions using as outcome 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 taking 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 (shown in Table 3.A5).

As a last exercise I create local school rankings inside each of the 26 states and federal district of Brazil, to 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 (Table 3.A6).

3.7 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 estimate the effect of disclosing information on school performance on enrollment decisions. 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 enrollments 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 face capacity constraints and thus prefer to adjust prices or select students based on ability rather than accept more students. Indeed Firpo, Ponczek and Possebom (2014) find that private schools in Brazil increased their fees following the disclosure of *Enem* school grades, although this occurred gradually over time. However, this would not explain why low-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 sub-population 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. (2017). Information, Market Incentives, and Student Performance: Evidence from a Regression Discontinuity Design in Brazil. *Journal of Human Resources, forthcoming.*

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.

Firpo, S., Ponczek, V., and Possebom, V. (2014). Private Education Market, Information on Test Scores and Tuition Practices. *IZA Discussion Paper (8476).*

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. (2009). 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*.

Tables and Figures

Table 3.1: 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%
Municipal schools	55%	0.9%
State schools	30.6%	84.9%
Federal schools	0.1%	1.5%

Table 3.2: 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 3.3: Summary statistics

	2005		2006		2007		2008	
	private	public	private	public	private	public	private	public
	Panel A - School characteristics							
No. of schools	6,091	15,946	5,841	15,439	5,846	17,259	6,812	18,223
No. of schools $<$ 10 Enem takers	1,764	1,891	1,694	1,831	1,898	2,604	2,071	2,704
Avg. grade (out of 100)	55	42	52	41	62	48	57	45
Avg. enrollment in 1^{st} year of H.S.	101	248	100	234	51	182	51	173
Avg. enrollment in 2^{nd} year of H.S.	104	192	100	180	47	138	47	131
Avg. enrollment in 3^{rd} year of H.S.	123	163	112	155	46	117	46	112
% eligible students taking Enem	48	42	51	44	61	46	66	49
	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
% father w/ high school	55	13	57	14	61	15	58	15
% mother w/ high school	62	18	64	19	68	21	65	21
Avg. age	17	19	18	21	18	21	18	20

*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 250 to 460 USD at the time approximately).

Table 3.4: 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	48	48	48	48	47
% Low income*	41	42	43	44	46	47
% Black	7	7	7	7	8	8
Avg. enrollment in 1^{st} year of H.S.	63	60	72	87	104	233
% of eligible students taking Enem	22	40	46	48	48	53
% Private schools	45	44	40	36	30	16
% Rural schools	13	11	8	6	5	1
Total obs.	9,131	9,174	8,729	7,861	6,738	49,824

*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 250 to 460 USD at the time approximately).

Table 3.5: Jumps in covariates

	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		0.4		0.1		0.3	
Obs.	19,635		19,635		19,635		19,636	

	% 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	0.4		3.7		4.5	
Obs.	19,636		11,031		11,196	

Note: Controls are year dummies. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations. The definition of low income 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 250 to 460 USD at the time approximately). Standard errors clustered at the level of municipalities in parentheses.
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Table 3.6: 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
$t = 2005$					
Enem takers ≥10	22%	18%	47%	13%	100%
Enem takers <10	21%	25%	35%	19%	100%
$t = 2006$					
Enem takers ≥10	23%	23%	43%	11%	100%
Enem takers <10	18%	25%	38%	19%	100%
$t = 2007$					
Enem takers ≥10	23%	17%	55%	5%	100%
Enem takers <10	22%	26%	42%	10%	100%

Note: 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 3.7: Main results

Cutoff rule	50 mark (1)	50 mark (2)	40% top/bottom schools (3)	40% top/bottom schools (4)	20% top/bottom schools (5)	20% top/bottom schools (6)
			Best-performing schools			
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)
Obs.	3,194	3,194	1,706	1,706	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)
Obs.	873	873	2,493	2,493	1,305	1,305
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)
Obs.	373	373	1015	1,015	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)
Obs.	5,774	5,774	2,969	2,969	1,715	1,715
Quadratic	No	Yes	No	Yes	No	Yes

Note: The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in year t. In columns 1-2 the best and worst schools are separated using the 50 mark cutoff in a scale of 0-100. In columns 3-4 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 respectively. In columns 5-6 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 clustered at the level of municipalities in parentheses.

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3.8: Estimates including measures of competition

Measure of competition	HHI schools (1)	HHI schools (2)	% private schools (3)	% private schools (4)	HHI grades (5)	HHI grades (6)
	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 ($Comp_{it} * D_{it}$)	0.005	0.005	-0.610	-0.618	-0.535**	-0.542**
	(0.07)	(0.07)	(0.27)	(0.28)	(0.23)	(0.23)
Obs.	3,194	3,194	3,194	3,194	3,194	3,194
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 ($Comp_{it} * D_{it}$)	-0.040	-0.042	0.059	0.070	0.078	0.090
	(0.09)	(0.09)	(0.34)	(0.35)	(0.29)	(0.30)
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 ($Comp_{it} * D_{it}$)	-0.124	-0.143	0.65	0.695	-0.298	-0.296
	(0.22)	(0.22)	(0.80)	(0.81)	(0.67)	(0.69)
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 ($Comp_{it} * D_{it}$)	-0.013	-0.012	0.317	0.314	0.215	0.213
	(0.04)	(0.04)	(0.25)	(0.25)	(0.20)	(0.20)
Obs.	5,774	5,774	5,774	5,774	5,774	5,774
Quadratic	No	Yes	No	Yes	No	Yes

Note: The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in year t. All regressions use the baseline specification where the 50 mark cutoff in a scale of 0-100 is used. In columns 1-2 the measure of competition controlled for is the HHI of school market shares in the municipality. In columns 3-4, the competition variable used is the share of high school students enrolled in private schools in the municipality. In column 5-6 the competition variable used is the HHI of the dispersion of school grades in the municipality. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations. Standard errors clustered at the level of municipalities in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3.9: Estimates including socio-economic factors

Socio-economic variable	Gini coefficient		Income per capita		Years of education	
	(1)	(2)	(3)	(4)	(5)	(6)
			Best-performing schools			
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 ($SE_{it} * D_{it}$)	-0.003	-0.003	-0.000	-0.000	-0.006	-0.006
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Obs.	3,194	3,194	3,190	3,190	3,190	3,190
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 ($SE_{it} * D_{it}$)	0.003	0.004	-0.000	-0.000	-0.006	-0.006
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)
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 ($SE_{it} * D_{it}$)	-0.011	-0.011	0.000	0.000	0.046	0.048
	((0.01)	(0.01)	(0.00)	(0.00)	(0.03)	(0.03)
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 ($SE_{it} * D_{it}$)	0.002	0.002	0.000	0.000	0.003	0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Obs.	5,774	5,774	5,743	5,743	5,743	5,743
Quadratic	No	Yes	No	Yes	No	Yes

Note: The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in year t. All regressions use the baseline specification where the 50 mark cutoff in a scale of 0-100 is used. In columns 1-2 the socio-economic variable controlled for is the Gini coefficient of the municipality in 2000. In columns 3-4, the variable used is the income per capita of the municipality. In columns 5-6, the variable used is average years of education in the municipality. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations. Standard errors clustered at the level of municipalities in parentheses.

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Figure 3.1: 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).

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

ESPECIAL
Tudo sobre o ENEM

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

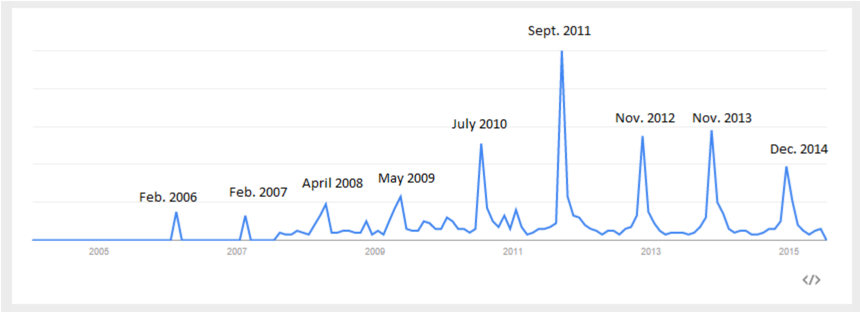
LEGENDA

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 3.2: Evolution of Internet searches on *Enem* rankings (Google 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 3.3: Frequency of schools by number of *Enem* takers

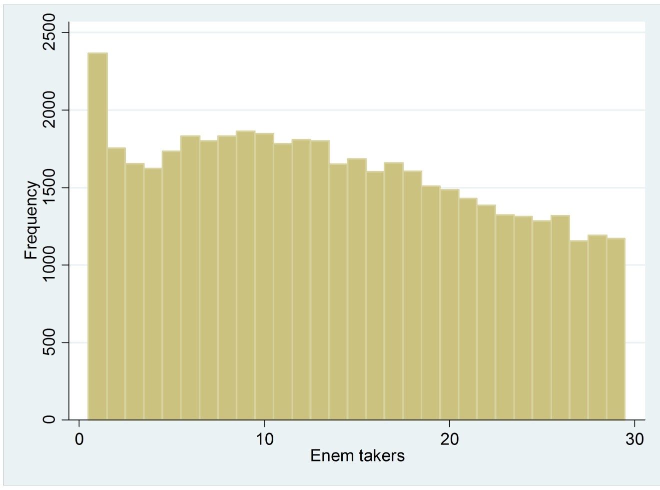
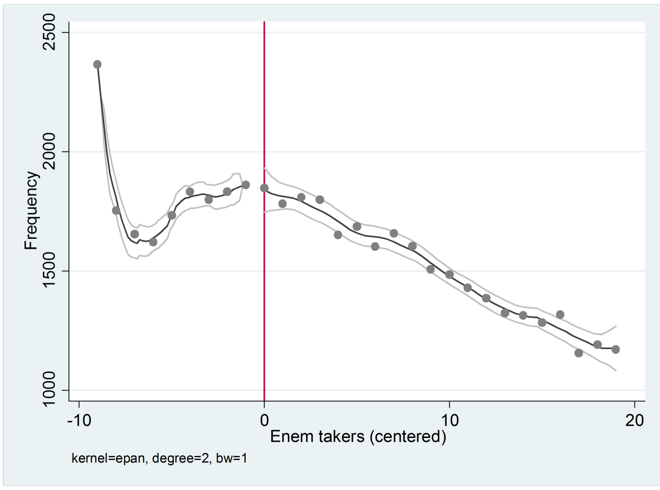


Figure 3.4: Local polynomial fit of frequency of schools by number of *Enem* takers



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

Figure 3.5: Distribution of test scores of public and private schools

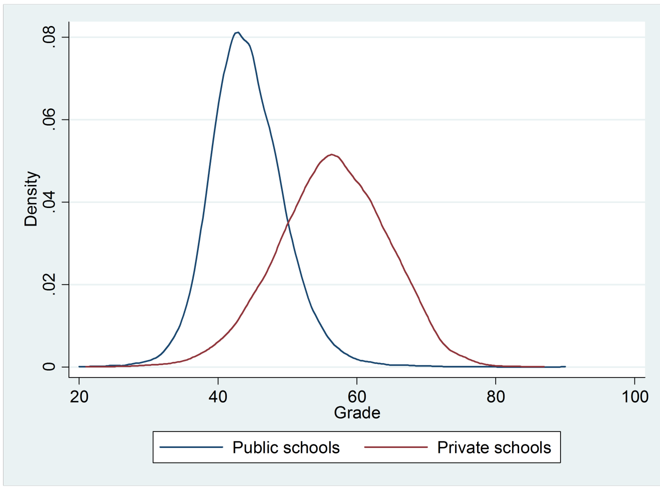
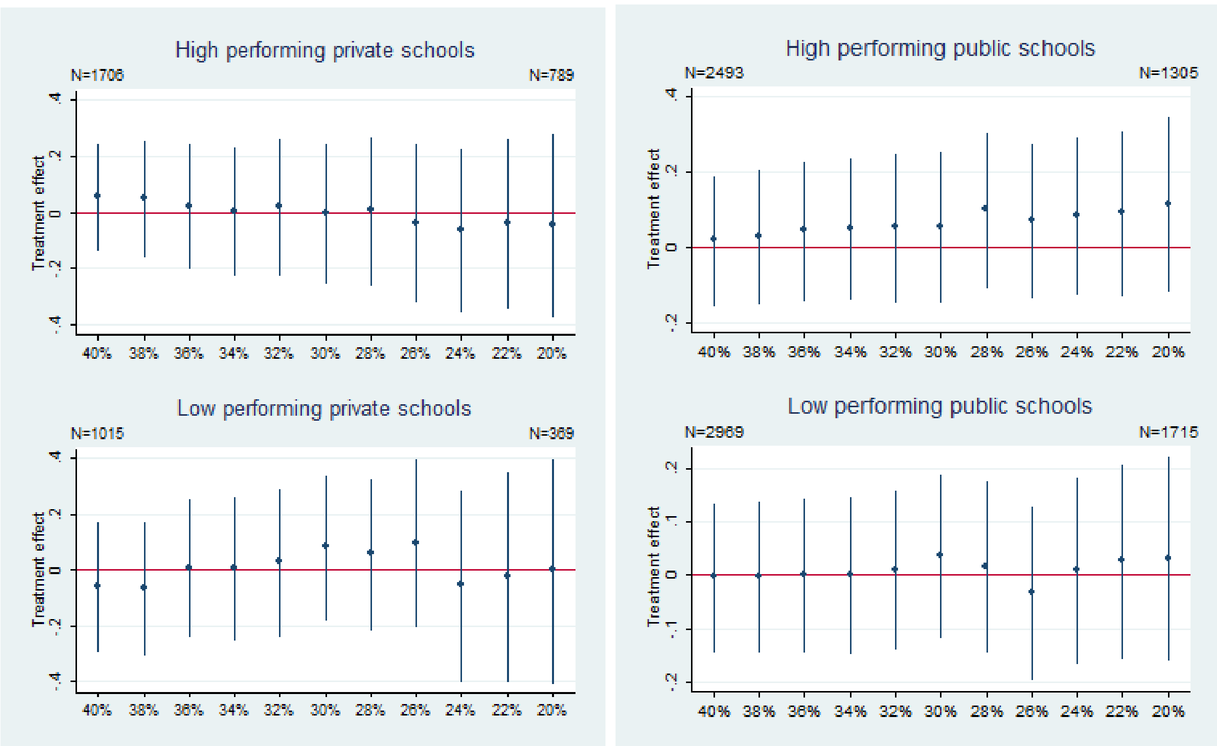


Figure 3.6: Plot of treatment effects using different cutoffs



Note: These graphs represent dummy coefficient values obtained in regressions using different cutoffs for defining the sample of high- and low-performing schools. For the best-performing schools (above), the first point shows the coefficient obtained when considering the 40% best schools, and the last point shows the coefficient obtained when considering the 20% best schools. For the worst-performing schools (below), the first point shows the coefficient obtained when considering the 40% worst schools, and the last point shows the coefficient obtained when considering the 20% worst schools. All estimations include a quadratic term for the number of test takers.

Appendix to Chapter 1

Table 3.A1: Jumps in covariates - 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		0.2		0.1		0.5	
Obs.	8,352		8,352		8,352		8,352	

	% 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	0.6		3.0		4.0	
Obs.	8,352		4,031		4,117	

Note: Controls are year dummies. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations. The definition of low income 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 250 to 460 USD at the time approximately). Standard errors clustered at the level of municipalities in parentheses.
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Table 3.A2: Jumps in covariates - 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		0.6		0.1		0.1	
Obs.	11,283		11,284		11,284		11,284	

	% 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	0.1		4.1		4.9	
Obs.	11,283		7,000		4,117	

Note: Controls are year dummies. A symmetrical window of data around the discontinuity of 5-15 Enem takers is used in the estimations. The definition of low income 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 250 to 460 USD at the time approximately). Standard errors clustered at the level of municipalities in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3.A3: Robustness test - window of 4-16 *Enem* takers

Cutoff rule	50 mark		40% top/bottom schools		20% top/bottom schools	
	(1)	(2)	(3)	(4)	(5)	(6)
			Best-performing schools			
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)
Obs.	3,703	3,703	1,975	1,975	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)
Obs.	997	997	2,866	2,866	1,497	1,497
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)
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)
Obs.	6,676	6,676	3,432	3,432	1,987	1,987
Quadratic	No	Yes	No	Yes	No	Yes

Note: The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in year t. In columns 1-2 the best and worst schools are separated using the 50 mark cutoff in a scale of 0-100. In columns 3-4 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 columns 5-6 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. Standard errors clustered at the level of municipalities in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3.A4: Robustness test - window of 3-17 *Enem* takers

Cutoff rule	50 mark		40% top/bottom schools		20% top/bottom schools	
	(1)	(2)	(3)	(4)	(5)	(6)
			Best-performing schools			
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)
Obs.	4,172	4,172	2,250	2,250	1,057	1,057
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)
Obs.	1,157	1,157	3,298	3,298	1,734	1,734
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)
Obs.	515	515	1,359	1,359	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)
Obs.	7,663	7,663	3,927	3,927	2,266	2,266
Quadratic	No	Yes	No	Yes	No	Yes

Note: The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in year t. In columns 1-2 the best and worst schools are separated using the 50 mark cutoff in a scale of 0-100. In columns 3-4 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 columns 5-6 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. Standard errors clustered at the level of municipalities in parentheses.
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Table 3.A5: Robustness test - total high school enrollment as outcome variable

Cutoff rule	50 mark		40% top/bottom schools		20% top/bottom schools	
	(1)	(2)	(3)	(4)	(5)	(6)
			Best-performing schools			
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)
Obs.	3,253	3,253	1,737	1,737	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)
Obs.	883	883	2,515	2,515	1,318	1,318
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)
Obs.	389	389	1,049	1,049	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)
Obs.	5,837	5,837	2,999	2,999	1,730	1,730
Quadratic	No	Yes	No	Yes	No	Yes

Note: The dependent variable is the log of total enrollments in high school. Controls are year dummies and enrollment data in year t. In columns 1-2 the best and worst schools are separated using the 50 mark cutoff in a scale of 0-100. In columns 3-4 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 columns 5-6 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 clustered at the level of municipalities in parentheses.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3.A6: Robustness test - State local rankings

Cutoff rule	50% top/bottom schools	
	(1)	(2)
	Best-performing schools	
Treatment effect - Private	0.001	0.018
	(0.04)	(0.06)
Obs.	3,441	3,441
Quadratic	No	Yes
Treatment effect - Public	0.028	-0.053
	(0.05)	(0.09)
Obs.	2,376	2,376
Quadratic	No	Yes
	Worst-performing schools	
Treatment effect - Private	-0.013	-0.068
	(0.26)	(0.42)
Obs.	126	126
Quadratic	No	Yes
Treatment effect - Public	-0.018	-0.003
	(0.03)	(0.06)
Obs.	4,271	4,271
Quadratic	No	Yes

Note: The dependent variable is the log of enrollments in the 1st year of high school. Controls are year dummies and enrollment data in year t. 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.

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$