UNIVERSITY OF SUSSEX

Essays on the impact evaluation of education policies in Mexico

by

Francisco Javier Cabrera Hernández

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy in Economics

in the

Department of Economics

August 2016

Declaration of Authorship

I, FRANCISCO JAVIER CABRERA HERNÁNDEZ, declare that this thesis titled, 'ESSAYS ON THE IMPACT EVALUATION OF EDUCATION POLICIES IN MEXICO' and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this University.
- I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____

Date: ____

"Reason is given to us as a practical faculty, that is, one that is meant to have an influence on the will. If nature has everywhere distributed capacities suitable to the functions they are to perform, reason and it alone, can produce a will that is good in itself."

Immanuel Kant

UNIVERSITY OF SUSSEX

Doctor of Philosophy in Economics

"Essays on the impact evaluation of education policies in Mexico"

Abstract

by Francisco Javier Cabrera Hernández

This thesis gathers research on three impact evaluations of interventions at the school and student level in Mexico. The first chapter evaluates the effects of a School Breakfast Program (SBP) on children's outcomes such as cognitive skills, illness, height and weight and grade repetition in the period 2002 to 2005. Quasi-experimental estimations provide evidence of positive effects on children's weight; however, such gains push children over their 'ideal' standardized average causing them overweight. This effect is significantly higher in the case of poorer children. The second chapter evaluates a Full-Time Primary Schools Program implemented in 2007, to work out if changing the time pupils spend at school can enhance skills in language and mathematics. Differences in Differences regressions point to a significant improvement of 0.11 standard deviations in mathematics and Spanish test scores after four years of treatment. These gains are three times higher in schools located in deprived areas and do not seem to be driven by students self-selection. The last chapter focuses on an exogenous policy change in Mexico which eliminates enforced grade repetition for all first to third grade students. This reform helped schools to reduce repetition rates from varying higher levels to almost zero in one academic year. Estimations coming from two-way fixed effects models using a panel of schools show an average reduction in dropout rates after reform implementation of 0.3% points along with no seeming effects on pupil's performance. General findings from the three chapters are of strong significance when placed into the broader debate about what works best in schools for improving children's academic performance and general education outcomes in Mexico.

Acknowledgements

I would like to express my appreciation and thanks to my supervisors Dr. RICHARD DICKENS and Dr. SHQIPONJA TELHAJ, for their valuable comments and patience. I would like to thank you for encouraging my research and for allowing me to grow as a scientist. I also want to thank other members of the Department of Economics at the University of Sussex, Dr. IFTIKHAR HUSSAIN and Dr. PEDRO ROSA, for going beyond your obligations and dedicate time to comment on my research and always make brilliant suggestions. Your advice on research, teaching as well as on my career has been priceless. I also would like to express my gratitude to Dr. SERGIO CÁRDENAS who has helped me to start my career as an academic in Mexico, Dr. CRISTIAN BELLEI who introduced me to the fascinating and challenging world of the educational policies and to the National Council of Science and Technology (CONACYT) for fully funding my research degree.

A special thanks to my family. Words cannot express how grateful I am to my parents VICTOR CABRERA and GLORIA HERNÁNDEZ, my brothers CARLOS and CÉSAR and my sister ALMA, I received the best from all of you. Thank you for showing me that with discipline, effort and good sentiments, all dreams can be achieved. Thank you for your counselling and the sacrifices that you have made on my behalf. I also want to acknowledge my sisters and brother in-law MAGDALENA, FABIOLA and FERNANDO for raising my beautiful nieces and nephews to whom I dedicate my research, PAMELA, ZAIRA, ALEJANDRO, PAOLA, ANDREA, CÉSAR and VALERIA. I hope this inspires you to achieve all your goals.

I would also like to thank my friends in the UK, who supported me and encouraged me to strive towards my goal, GUSTAVO IRIARTE, EVA EGGER, Dr. FRANCESCO DI BERNARDO, Dr. MIMI XIAO, SEBASTIAN WELISIEJKO, RENEEMA HAZARIKA, TAKAHIRO KIDA, DAVID MARTINEZ and Dr. PEDRO ORRACA. I also want to thank my old friends in Mexico, who in the distance kept me in their thoughts, ERNESTO CASTRO, JAZMÍN PALACIOS, JESSICA VENEGAS, CITLALI CANSINO, ARMANDO SUÁREZ and ALEJANDRO ARVÍZU.

Last but not least, I would like to express appreciation to my beloved partner and best friend, ADRIANA, for her love, encouragement and assistance throughout writing this thesis.

Contents

List of Figures

List of Tables

To my parents. . .

Introduction

In recent years Mexico has transited through a period of high flux in terms of education policies. The last two Mexican presidents (2006-2012 and 2012-2018), for example, have introduced the so-called ”Educational Reform” as a key topic in their political agendas and public speech. The first step towards placing education at the center of the debate included the creation in 2006 of a massive national standardised evaluation of pupil’s knowledge in mathematics, Spanish and science to diagnose the state of the education in Mexico, it was named the ENLACE test.

Results were not promising. From the first waves of the national standardised test, estimations showed that around 70% of Mexican children had an ‘insufficient’ level of knowledge. International tests, such as PISA and TIMMS, only confirmed that Mexico was among the worst in the Organization for Economic Cooperation and Development (OECD). For example, the 2009 PISA results ranked Mexico as the 48th place in reading and 50th in math out of 65 evaluated countries and furthermore, they also revealed that the country was not performing better than others with similar income, such as Uruguay and Chile and even poorer countries like Cuba.

The results coming from the national and international evaluation were worrying as economic research has provided evidence that improvements on math, language and science test scores relate to increases in economic growth (Hanushek and Kimko, 2000; Barro and Lee, 2001; Hanushek, 2013), earnings in adulthood (Murnane, Willett, and Levy, 1995; Murnane, Willett, Duhaldeborde, and Tyler, 2000; Lazear, 2003) and to the reduction of the inequality of income between social groups (Hanushek, 2004). Evidence has also shown that there are other non-monetary benefits from education such as improved health status and lowered crime (Lochner, 2011). However the literature is still limited on providing a clear guidance of what policies and specific investments should be pursued to increase educational outcomes in specific contexts and thus, it has proved to be harder for governments to design policies based on conclusive evidence.

The Mexican government opted to start a range of different policies addressing diverse problems in basic schools seeking to close the knowledge gaps with other similar and developed countries. Nowadays a total of nine policies are present at the school level in basic education. The most relevant programs due to their size and public investment address diverse challenges and topics, ranging from health and safety interventions "Escuelas Seguras", infrastructure "Escuelas Dignas", information technology "Programa de Inclusion Digital", to programs seeking to decentralise schools' decisions and increase their autonomy ”Escuelas de Calidad", to help children's nutrition through the provision of free-breakfast ”Desayunos Escolares del DIF", and to lengthen the school day ”Programa Escuelas de Tiempo Completo". Further reforms have sought to change the structure of teacher's evaluations and their monetary incentives ”Carrera Magisterial" along with the way children are assessed in order to facilitate their promotion to higher education levels.

Along with the supply in education programs for basic education, federal and local governments spending on education has also increased. According to the OECD[1] Mexico's spending on education as a proportion of its GDP grew from 4.4% to 5.5% in 2014. The amount spent on primary and secondary education (3.9%) is even higher than that of the OECD countries (3.7%). Between 2005 and 2012 only, the money invested in primary and secondary schools observed a real growth of 11% per student. However, despite the increasing efforts to design and finance a broad range of policies, most of the programs implemented have not been evaluated, for which it remains unknown whether the money spent on these, financed through taxation, has improved general education outcomes or not.

This thesis seeks to contribute to education policy in Mexico by presenting three impact evaluations. The first one examines one of the oldest policy interventions in this country, the school's breakfast program "Desayunos Escolares del DIF". The second and third one relate to two recent modifications seeking to increase the time that pupils spend in primary schools by lengthening the school day from four and a half to eight hours daily (”Programa escuelas de tiempo completo" or PETC, for its abbreviation in Spanish), and a recent law change that abolished grade retention during the first three grades of primary education seeking to reduce dropout rates among Mexican children of age 6 to 9. These three policies are representative of the different strategies adopted by the Mexican government as they differ in their organisation objectives and costs.

[1] See the 2015 ”Education at a Glance" report

The first chapter of this thesis includes the impact evaluation of the school breakfast program. Starting in the 1960s, the program is still active and it has not been broadly updated since. As it will be explained in detail, its presence is now often related to political reasons rather than to school improvements. Nonetheless, only between 2007 and 2012, approximately 3.8 billion dollars were invested in the distribution of about 3.5 billion breakfasts nationally, accounting for approximately 0.05% of the annual Gross Domestic Product (GDP) in that period.

At the moment this evaluation was conducted. The school breakfast program was the second largest social program in Mexico. For example, in 2012, 12.2% of the households in Mexico had at least one member receiving a school breakfast, a number only below PROGRESA.[2] However, to the best of my knowledge, the evaluation presented in Chapter 1 is the first one on its potential causal impacts on health, cognitive outcomes and school performance at the national level of this large school breakfast program that is aimed to be an important component of the social safety net of children in Mexican low-income households.

The second chapter evaluates the effects of the full-time schools program PETC, on primary student's academic achievement, measured in standardised test scores of mathematics and Spanish from 2008 to 2013. PETC seeks to improve learning opportunities by increasing the time children spend at school from four and a half to eight hours everyday, while incorporating new subjects and activities in the curricula (e.g foreign languages, arts, culture and nutrition). The program grants every year a fixed stipend for operative expenses and a varying fund according to the number of professors and students in each school.

PETC started in 2007 and up to 2013 it represented a spending of about US$460 millions. Moreover, the 2012 elected federal government announced an expansion of the program in order to reach 40,000 primary and secondary schools by 2018. According to the Secretariat of Finance in Mexico, the 2014-2015 budget programmed for PETC rose to US$1 billion. Nevertheless, to the best of my knowledge, the most recent expansion of PETC was dictated without any previous public evaluation on the potential causal impacts on school and children's outcomes such as test scores and grade repetition at the national level of this large program that is aimed to be an important component of the educational strategy in Mexico.

[2] PROGRESA, OPORTUNIDADES or PROSPERA, as named since 2012, is a poverty reduction conditional transfer program that has important education and health components, as families receive transfers for sending their children to school and to visit health clinics (Rodríguez, 2005).

The third chapter presents the impact evaluation of abolishing grade retention in the first three grades of primary school on pupil's dropout rates and children's performance. In 2012, Mexico made an important change in its evaluation policies and passed from a punitive system in which all children graded below a given threshold were not promoted to the next academic level, to one focused on the mandatory "social promotion" of students. The reform allowed children in the first three years of primary education to be automatically promoted regardless of their level of achievement. This change caused an exogenous reduction of grade repetition rates independent of school's characteristics and previous trends.

This change in the retention laws varies from the first two programs analysed in Chapters 1 and 2 as it did not imply a modification of school's inputs and consequently it did not represent any increase in resources. Therefore, the chapter includes a broader discussion on the determinants of dropout rates and how this can vary independent of children's socioeconomic status and family investments in the context of a low-cost reform. Interestingly, maybe because the policy change did not imply a big initial investment, the rules to promote young primary students are now back to what they were before 2012, as a new change was introduced in 2012 that brought back the previous punitive system of evaluation and promotion for children in second and third grade. Similar to the original reform, the counter-reform took place without any empirical evidence of what could the be positive or negative effects attached to it.

The main conclusions coming from the different empirical analysis in this study represent useful inputs to redirect or reinforce education policy efforts in Mexico and offer a clearer idea on how to improve the specif policies addressed. For example, effects on the school breakfast program suggest that it is causing overweight problems among Mexican children, specially among the poorest. This exhibits the importance of updating the objectives of such policy. Similarly, as it is thoroughly discussed in Chapter 2, lengthening the school day seems to be having a positive effect on children's achievement, and this positive result is three times higher in more deprived schools. These outcomes works to justify a more informed expansion of the program. Last but not least, Chapter 3 offers some support to social promotion as it shows a robust reduction on dropout rates, and offers a guidance for future amendments to retention laws in primary schools.

Chapter 1

Impacts of the School Breakfast Program in Mexico on Children's outcomes: IQ, Health and Grade Repetition

Research in laboratories has found negative effects of short-term fasting on cognitive abilities like short term memory and attention due to low glucose concentrations, particularly among children who are at nutritional risk.[1] Plausibly, this may occur due to the short-term absence of nutrients and glucose needed for the cognitive processes demanded in the learning process (Pollitt, 1995). Moreover, the sustained absence of breakfast can also have an impact on children's physical development in the long term, affecting their immune system and their general health (Pollitt, 1995). In sum, the long absence of nutrients may result in sickness and, by consequence, in lower school attendance which may potentially affect children's educational outcomes.

Considering the plausible detrimental effects of breakfast absence on education outcomes, different school breakfast or lunch programs have been applied at the school level in developed and developing countries seeking to improve the nutritional status of children. These kind of polices have been generally directed towards those who are undernourished or come from an impoverished background, where the risk of a poor nutrition is higher. Plausibly, the provision of breakfast in schools can potentially increase low-income children's nutritional status and improve their academic outcomes.

[1] For a review of these studies see Pollitt, Cueto, and Jacoby (1998)

A policy of free breakfasts could also make poorer children healthier and decrease absenteeism due to illness. Finally, the provision of free food for deprived children could also represent money savings for poor families allowing them to reallocate resources towards the acquisition of other inputs related to offspring's education.[2]

In regard to this, a few studies based on Randomized Control Trials (*RCT*) have shown that the provision of breakfast has had positive effects on different outcomes like height and weight, daily attendance and the reduction of drop-out rates in Jamaica (Powell, Walker, Chang, and Grantham-McGregor, 1998), Peru (Cueto and Chinen, 2008), United Kingdom (Shemilt, Harvey, Shepstone, Swift, Reading, Mugford, Belderson, Norris, Thoburn, and Robinson, 2004), and in other developing countries[3]

Notwithstanding, more recent evidence for the United States (US) have brought into consideration the possible negative general effects of a free food or reduced price policy on children's health. For example, Schanzenbach (2009) studies the effect of the National School Lunch Program (NSLP) in the US taking advantage of a discontinuity in the eligibility for a reduced price lunch. The author found that the obesity rate in participants is higher than in nonparticipants. In a second study, Millimet, Tchernis, and Husain (2010) using a Propensity Score Matching (PSM) approach, also found a positive impact on weight concluding that the NSLP contributes to obesity. However, the authors also analyse the effect of the School Breakfast Program (SBP) in the US and found that this, by the contrary, reduces overweight. As argued by Hinrichs (2010), one of the main problems in the evaluations conducted in the US is the presence of two similar programs, the NSLP and SBP, and the confounding effects coming from these.

This chapter focuses on the evaluation of the School Breakfast Program in Mexico, and its potential effects on cognitive outcomes, height and weight by sex and age, illness and grade repetition probabilities of children in a sample extracted from two waves of the Mexican Life and Family Survey (MxFLS) for the years 2002 and 2005.[4] To the best of our knowledge this is the first paper to evaluate the potential causal impacts on health, cognitive outcomes and school performance at the national level of this large child nutrition program that is aimed to be an important component of the social safety net of children in Mexican low-income households.

[2]For example, Babu and Hallam (1989) found that the provision of food from school feeding programs enables households to increase expenditures on non-food goods.

[3]A review of the literature finding positive effects on attendance rates in Bangladesh, Haiti, India, Brazil and Honduras can be also found in CARE (2004).

[4]A detailed description of the data is offered in Section 1.2.

Different nutrition programs directed to vulnerable people in Mexico have been implemented since 1929, when free rations of milk started to be offered in deprived schools of Mexico City under the program "A Drop of Milk". Since then, with the creation of the National Association for the Infant Protection (INPI, for its abbreviation in Spanish) in the 1960s, modified versions of a breakfast program became the core of the alimentary strategy at the national level.[5]

From 2007 to 2012, approximately 3.8 billion dollars were invested in the distribution of about 3,5 billion breakfasts nationally, accounting for approximately 0.05% of the annual Gross Domestic Product (GDP) in that period. The SBP is the second social program with highest number of beneficiaries in Mexico. In 2012, 12.2% of the households in Mexico had at least one member receiving a breakfast, a number only below PROGRESA,[6] the most important social program in Mexico, that in the same year reached 18.8% of the households with a cost of 30.8 billion dollars or 0.4% of the GDP (Gutierrez, Rivera, and Shamah, 2012). However, as mentioned, there has not been an evaluation of the SBP effects at the national level.

The 32 state governments in Mexico publish a yearly administrative evaluation of the SBP.[7] However, these documents do not try to find any causal effects on specific outcomes and rely exclusively on the description of costs, implementation issues, the amount of children treated and in exceptional cases the description of the socioeconomic characteristics and nutritional habits of the children in the program.[8] Although these studies are relevant to understand the application and targeting of the program, they do not offer further conclusions about its potential effects.

Other evaluations seeking to find causal effects of different breakfast interventions in Mexico are inconclusive.Vera (2005) performs an evaluation of a breakfast intervention for preschool children on cognitive-motor development of preschool students (4-5 years old) in deprived and rural areas in Sonora, a northern state of Mexico. The author evaluated different attention, memory and cognitive skills before and after the program among 300 children treated and 150 untreated. The results show positive effects on

[5] All historical references as well as the figures regarding government spending on the breakfast program are taken from SNDIF (2012).

[6] PROGRESA, OPORTUNIDADES or PROSPERA, as named since 2012, is a poverty reduction, conditional transfer program that has important education and health components, as poorer families receive transfers for sending their children to school and to visit health clinics (Rodríguez, 2005). PROSPERA scholarships are thus assigned to the poorest children in each school, according to a socioeconomic survey applied to households located in identified vulnerable areas in Mexico.

[7] This represents 31 states conforming the Mexican Federation and Mexico City

[8] See for example, Shamah, Morales, and Hernandez (2010); DIF (2011); SNDIF (2012)

attention and perception processes of the children under treatment. However, authors picked up schools that could not be treated for political and administrative reasons as their counter factual. One problem with this setup is that such political and administrative issues in this case related to school's inputs and pupil's socioeconomic environment, for which the "experimental" design did not achieve the comparability of the socioeconomic characteristics of the control and treatment groups, specifically regarding income and goods' availability. Moreover, the methodology presented does not try to control for such differences. This undermines the validity of the results.

Similarly, Ramirez, Grijalva, Valencia, Ponce, and Artalejo (2005) attempt to evaluate the SBP's impact on weight and other bio-chemical indicators of vascular risk in 17 municipalities of Sonora. The authors compared the results of 254 children treated and 106 untreated after one academic cycle. The results do not show any significant effects on neither, Body Mass Index (BMI), the percentage of fat, cholesterol, sugar levels nor cardiovascular risk. However, this study suffers form a high attrition rate, since approximately 40% of the students treated dropped out from the experiment and does not discuss how this issue may have affected their results given that, apparently, the most deprived children are the ones drooping out from their study.

Assessing the causal effects of a program like SBP is difficult for two main reasons. First, if the program is focused on disadvantaged children, those receiving free breakfasts are likely to differ from non-participants in observed and unobserved ways not present in the data.[9] Secondly, there is an apparent threat from the under-report of participation in programs such as the SBP, that may bias the estimations. In this regard, Gundersen, Kreider, and Pepper (2012) argue that miss-reporting could be one of the reasons why some of the evaluations of the NSLP in the US have found negative effects on children's obesity.

When subjects self-select into the treatment group on the basis of attributes unobserved by the researcher, but correlated with the outcome of interest, the estimations are no longer suitable to find causal effects. However, as it will be presented in Section 1.4, the identification strategy of this paper relies on the fact that the assignment of the program is in practice determined by school variables which are observable (i.e. principals' education and gender, number of teachers and number of students) rather than children's

[9] For example Millimet and Tchernis (2012) show that unobserved ways of self-selection into treatment are important in the results of the evaluation of the SBP in the US. The authors conclude that the SBP has a positive *relation* with children's weight in estimations that do not control for unobserved heterogeneity, nonetheless, when the authors use an IV estimator that exploits heteroscedasticity for identification, they found a *causal* negative effect on weight.

characteristics such as height and weight and/or household's or parent's characteristics. This allows for a statistical matching with a broad *common support*, useful to compare the outcomes of children with similar background characteristics who are treated and untreated.Therefore, in addition to OLS estimations I present the results of a PSM using different parametric and non-parametric algorithms on a sample of children between 6 and 12 years old, in order to find the arguably causal effects of the SBP. To the best of my knowledge this is the first study to evaluate the SBP at the national level and also differentiates from Ramirez et al. (2005) by contributing to the analysis of heterogeneous effects among poorer children and measures on educational attainment and cognitive outcomes.

Another important characteristic of this study is that program participation does not come from self-reporting but from administrative records. The data used includes a section with information about the schools in the municipalities where the survey was conducted, allowing to identify those schools that are participating in the breakfast program. The data also permits us to relate children and background characteristics, like their parent's education and occupation to the schools in which they study. In effect, underreporting should not be a major problem.

The results of this paper are generally similar in the OLS and PSM and despite some differences in their magnitude, they offer the same general conclusions. The provision of free breakfasts in primary schools in Mexico seems to have a positive effect on the weight of students between 9 and 12 years old, who have been treated for at least 4 and up to 6 years. This effect is stronger for those at the bottom part of the income distribution pushing them above their standardized average by age and sex. However there is no evidence of impacts on cognitive skills, health, children's height nor grade repetition.

The rest of this chapter is structured as follows. Section 1.1 includes a detailed description of The Integral Strategy for Social and Alimentary Assistance (EIASA, for its abbreviation in Spanish) in Mexico, which includes the SBP. Section 1.2 describes the data and the variables to be used and shows some descriptive statistics of the sample. Section 1.3 shows the main results obtained with OLS. Section 1.4 discusses the characteristics of the assignment into treatment. Section 1.5 exhibits the main impacts of the program on the outcomes described. Section 1.6 offers some concluding remarks and policy implications.

1.1 The Breakfast Program in Mexico

The EIASA seeks to improve the nutritional status of the vulnerable population in the country (i.e. elder, poor, homeless or children with low weight and height) and aims to promote better nutrition habits in order to avoid obesity among the population. It is managed by the National System for Integral Family Development (SNDIF, for its abbreviation in Spanish) a sub-office of the Ministry of Health in Mexico, and it consists of four different programs:[10]

- School breakfast: consisting of a daily breakfast for children in public primary schools in need of social assistance (i.e. with low height and weight)
- Alimentary assistance for children under 5 years old: it provides food to children who are not in school and in need of social assistance (i.e. with low height and weight)
- Alimentary assistance for vulnerable individuals: consisting of the distribution of a regular larder to elder, handicapped, pregnant or lactation women.
- Alimentary assistance for homeless families: it consists of the distribution of breakfast and meals.

As mentioned, this research is focused only on the evaluation of the School Breakfast Program. This policy operates with the use of national guidelines provided by the Federal government concerning the attributions and obligations of each of the 32 states of Mexico and defining the targeting of schools and children in the program. This guidelines state that the program should focus on public kindergarten and primary schools with a higher proportion of undernourished children, or in risk of being undernourished. Preferably, treated schools should be located in rural zones or deprived urban areas, and/or should have a higher concentration of indigenous people. (SNDIF, 2012, p.3).

Regarding program's funding and targeting, each of the 32 Mexican states receive federal funds based on a formula that takes into consideration the number of schools and students from vulnerable areas. Afterwards, state governments acquire from the private sector the necessary inputs for the breakfasts and decide which schools to be treated,

[10] All the information about the organization, application and funding of the SBP as discussed in this section comes from SNDIF (2012)

normally, supported by a pre-selection made by the municipalities. Consequently, municipalities have a strong influence on the choice of schools to be treated according to their "good will" or independent criteria to apply the national guidelines (SNDIF, 2012, pp.10-11).

In Mexico there are some established mechanisms for which the Ministry of Finance enforces the local governments to inform the destination of all the resources received from the Federation, and plausibly, this is also true for the resources received for the application of the SBP. Nonetheless, there is no mechanism to enforce the targeting of the SBP. Finally, school's principals are not allowed to provide SBP to certain students and all children in each of the schools included in the program are treated, regardless of their individual nourishing conditions.

A recent qualitative study gathers information about SBP targeting in a sample of four representative states of Mexico (Soto and Lorenzo, 2008).[11]. The results exhibit that the extension of the program differs according to each state's capabilities and limitations (i.e. infrastructure, budget, distribution line, etc.). However, all local governments declare that the program is "always directed to schools with low-size children that are located in deprived zones.[12] This is also declared by other governments in the central part of Mexico (i.e. Estado de Mexico, Federal District) in some official documents (see Shamah et al., 2010; DIF, 2011).

The specific contents of the breakfast are relatively homogeneous and even if there is some variation from state to state according to their availability of fresh produce, they are aimed to cover the same nutritional requirements. As a general rule, a cold breakfast should include, 250 milliliters of milk; a minimum of 25 grams of seeds and oils, one fresh fruit from the region or at least 20 grams of dry fruit. The hot breakfast should include 250 milliliters of milk, and a meal made of vegetables, cereals and/or legumes and a piece of fresh fruit.

The information explaining how the program is assigned is of great relevance for the analysis conducted on this paper. It is clear that, if the breakfast is effectively directed towards all children who are most in need, it could be difficult to find a counter-factual to evaluate the effects of the program. However, if the way states' authorities assign the program to those more in need is not effective enough, if for example, they choose the schools which are easier to reach rather than the more deprived, there is a chance

[11]Chiapas, Jalisco, Veracruz and Baja California

[12]With the exception of the state of Chiapas in the southern part of Mexico, that in the recent years chose to apply the SBP "universally" and declares to have reached already 95% of the students

to find students that should have been treated but are not treated and vice versa. A further discussion regarding the assignation into treatment in the case of this program is presented in Section 4.

Finally, it is important to notice that the nutritional challenges in Mexico are now different from what they were years ago when the SBP started at the national level. Mexican children have passed, on average, from undernourishment to obesity in the past decades. One explanation of this is that the diet of many families have changed to include products rich in fats and calories and low in nutrients. In fact, only 1.6% of the Mexican children between 5 and 11 years old are undernourished, 13.6% have a low size, 19.8% have excess weight and 14.6% are obese (Gutierrez et al., 2012). Recently, the alimentary strategy of the federal government has changed and created other initiatives to reduce children's obesity;[13] however, no major changes have been implemented on the design, operation and application of the SBP.

1.2 Data and Descriptive Statistics

This study uses a panel of households extracted from the MXFLS for the years 2002 and 2005. The dataset includes information on family characteristics such as parent's education, household's income and number of children in the two periods. It also includes a set of Raven's tests applied to all the members of the family which are useful to measure cognitive skills independent of schooling (Raven, 1998). Other individual measures are included in both waves, such as anthropometric variables like height and weight.

The 2002 survey also includes information of the characteristics of schools. These include variables such as the number of students, the number of teachers and principals information including their gender and education. It also includes variables regarding schools sanitation such as cleaning personnel and bathroom facilities. Unfortunately, school's characteristics are not measured again in 2005. The implications of only having school's characteristics measured in 2002 on the design of the evaluation are discussed below, where I define the sample to be used.

Further detail of the outcome variables to be measured as well as the control variables and the waves of the MxFLS on which they are present is shown in Table 1.1.

[13] One example is the "Acuerdo Nacional para la Salud Alimentaria (ANSA)", designed to combat children's excess weight and obesity (Gutierrez et al., 2012, p.154).

TABLE 1.1: Description of variables and MxLFS waves in which they are present

Variable	Description	Wave	
		02	05
Outcomes:			
Cognitive test scores	Raven's cognitive test scores standardized by age	x	x
Illness	A dummy variable that takes the value of 1 if the kid had a disease in the last months and zero otherwise	x	x
Height	child's height standardized by age and sex	x	x
Weight	child's weight standardized by age and sex	x	x
Grade Repetition	A dummy that takes the value of 1 if the kid has repeated grade during his primary studies	x	x
A. Context Variables			
Log of Income	Natural logarithm of the Household's income	x	x
Household's Head Schooling	Years of Schooling of the household's head	x	x
Spouse's Schooling	Years of Schooling of the spouse	x	x
Number of siblings	Child's number of sibling	x	x
PROGRESA	A dummy that takes the value of 1 if the household receives conditional transfers or zero otherwise	x	x
Girl	A dummy that takes the value of 1 if the kid is a girl or zero otherwise	x	x
Age	Kid's age in years	x	x
B. School Characteristics			
Multiple Grade School	A dummy that takes the value of 1 if the school has students from different grades in the same classroom and zero otherwise	x	
# of teachers	Number of teachers working in the school	x	
# of students	Number of teachers registered in the school	x	
Principals education	A categorical value that takes the value of 1 if the principal's highest level of education is high school , 2 University and 3 postgraduate studies.	x	
Breakfast	A dummy that takes the value of 1 if the school is participating in the Breakfast Program or zero otherwise	x	
Principal's sex	A dummy that takes the value of 1 if the principal is a female and zero otherwise	x	
Principal has another job	A dummy that takes the value of 1 if the principal has a second job outside the primary school or zero otherwise	x	
Cleaning Personnel	A dummy that takes the value of 1 if the school has full-time cleaning personnel and zero otherwise	x	
Bathroom water/soap	A dummy that takes the value of 1 if the school's bathroom has pipe water and soap and zero otherwise	x	

This rich dataset allows to include an important number of observable characteristics relevant to both, the analysis of the effects of the breakfast program and the assignment into treatment. However, there is an important limitation of the data at hand. As shown in Table 1.1, schools' information included in the survey, for which is possible to know whether any school is participating in the SBP or not, is only present in the survey of the year 2002. Additionally, it is only possible to know, retrospectively, if the program was applied during the scholar cycle 2000/2001 and 2001/2002. Therefore, it is not possible to know for how many more years the schools have been applying the program and if they continued applying it after 2002.

These limitations offer two challenges: first, it is not possible to know whether the children who are studying higher grades in a primary school participating in the SBP in 2002, received the treatment during all the time they were registered in that school or if they received it only during the previous and the present year, as registered by the survey. However, since the survey was taken from May to August of 2002, the way I address this problem is by restricting the sample to those children who are 6 to 9 years old in 2002 and/or study the first to the third year in the primary school. The logic of this restriction is shown in Figure 1.1 and it can be explained as follows:

- If children in a treated school are 6 years old or are in their first year in the summer of 2002 (letter **a** in Figure 1.1), they started receiving breakfast on September 2001, when the scholar cycle started for them. Therefore, it can be assured that they have received at least one academic year of treatment (i.e. September to June).[14]
- If children are 9 years old or in their third year in the summer of 2002 (letter **b** in Figure 1.1) and the school only implemented the program during the present academic cycle, they have been treated for at least one year; but if the school started implementing the program since the cycle 2000/2001 they have received two years of treatment.
- If the school has had the program from before the cycle 2000/2001 (something that is not possible to know with the available data) the children who are 9 years old or in their third year (letter **b** in Figure 1.1) have potentially received a maximum of three years of treatment since they were 6 years old when started primary school (letter **c** in Figure 1.1)

Consequently, the sample of children from 6 to 9 years old has received a minimum of 1 year and up to 3 years of treatment.

The second challenge arises because it is not possible to know if the schools in the SBP continued applying the program after 2002. However, as mentioned, it is possible to know how many schools were applying the program in the cycle 2000/2001 and how many continued doing it in the next cycle: 2001/2002. According to the MxFLS, only 5 out of 92 schools dropped out of the program from one year to the other. Given this low rate of schools dropping from the SBP, I assume that the schools in the program remain in it once they have been selected into treatment.

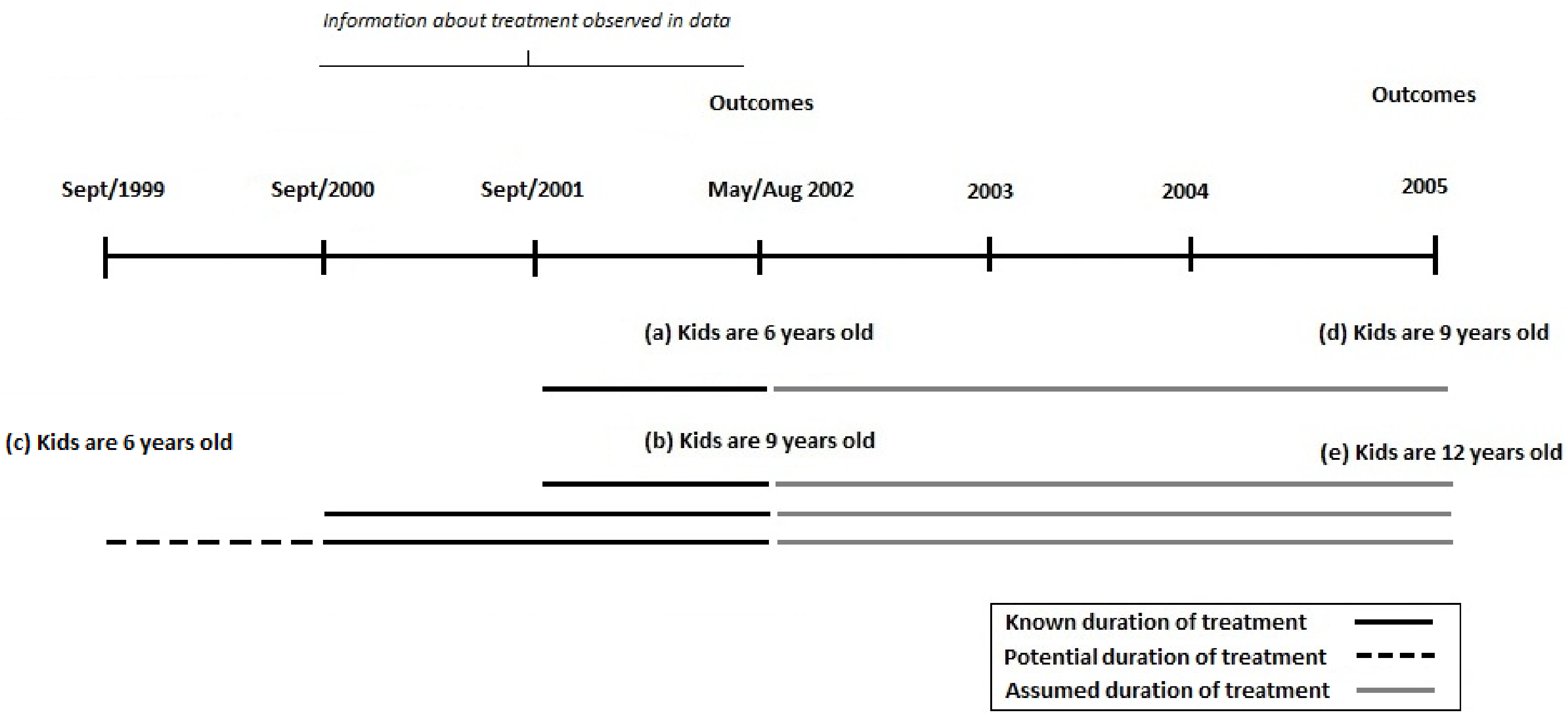
It is possible to "observe" in 2005 the same children who were receiving treatment in 2002 who are now between 9 and 12 years old and who are in the third to the sixth grade of their primary school studies (letters **d** and **e** in Figure 1.1). If the schools in which

[14] From now on I use year as an equivalent of academic year

they study remained in the program, these children have received at least four and up to six years of treatment (this is represented by the union of the dotted, the gray and the black line in Figure 1.1).[15]

[15] Approximately 20% of children received breakfast during 4 years and 80% during 5 and up to 6 years

FIGURE 1.1: Timeline:Observed, potential and assumed treatment duration for children 6 to 9 years old in 2002



Finally, the sample only includes children who attended the morning shift in public primary schools in Mexico, who did not stop attending school during the period 2002 to 2005 and who did not change their residency during that period. The sample of the 2002 cohort consists of 344 schools from which 94 were treated and 250 were untreated. In the balanced panel (2002-2005) there are 283 schools from which 87 are treated and 196 are not treated. This results in 1839 children for the cohort of 2002 and 1337 who can be observed in 2002 and 2005, as shown in Table 1.2.

TABLE 1.2: Sample of children 6 to 12 years old in cohorts and panel 2002-2005

Variable	Treated	%	Untreated	%	Total
Sample 2002	687	37	1152	63	1839
Panel 2002-2005	522	38	815	62	1337

The main regression framework includes four dependent variables: Children's Raven's scores, a dummy variable indicating if the child had a disease during the previous month and the standardized measures of height and weight by age and gender. These variables are regressed on a set of controls including context characteristics, individual attributes of children (i.e. height, weight and Raven's scores in 2002) and school characteristics including their sanitation. The same regressors are also used to analyze the selection into treatment and to calculate a Propensity Score to match students treated and untreated and obtain the Average Treatment on the Treated (ATT) and the Average Treatment Effect (ATE) of the SBP.

The regressions also include a dummy variable indicating whether the student is participating in PROGRESA or not. It is important to control for the presence of this program, because plausibly the SBP is directed to the same population than PROGRESA and the results here presented could be capturing the effects of such program and not the effect of the breakfast.

I use the data of the 2002 cohort to perform OLS regressions and PSM to observe the short-term effects of the SBP (i.e. 1 to 3 years of treatment). The same analysis is performed on the panel (2002-2005) to obtain the medium-term effects (i.e. 4 to 6 years of treatment).

Table 1.3 shows the main descriptive statistics of the variables used as controls in the OLS model and to calculate the propensity score in the matching process. It is possible to observe the differences between the treatment group and the untreated as well as their statistical significance. The main context variables that significantly differ between treated and untreated are the parent's years of schooling, the proportion of students receiving PROGRESA transfers and the standardized Raven's scores by age and sex. However, anthropometric measures such as height and weight and their standardized values are not significantly different between the two groups, despite being one of the main selection criteria declared by the local authorities. By the contrary, treated and untreated schools exhibit different characteristics.

For example, the proportion of children attending multiple grade schools is higher in the treated group.[16] The literature shows that in multiple grade schools there is more poverty, pedagogical difficulties and lower achievement than in full-grade schools (Cueto and Chinen, 2008, p.134). This is an intuitive result and is also consistent with the lower number of teachers and students on average in the treatment group. However, contrary to the intuition, principals' education in treated schools is significantly higher than in the untreated, and these teachers could be more prone to be located in better schools and urban areas. Furthermore, schools infrastructure, captured by the presence of bathroom facilities with tap water and cleaning products is higher on average for the treated group.

TABLE 1.3: Descriptive statistics of a panel of kids 6 to 9 years old in 2002

	Treated			**Untreated**			**Difference**	
	Obs.	Mean	S.D.	Obs.	Mean	S. D.		
A. Context Variables								
Log of Income	433	9.78	1.02	658	9.85	1.20	-0.07	
HH Years of Schooling	438	7.15	3.01	659	7.50	3.18	-0.35	*
Spouse Years of Schooling	407	7.06	2.85	561	7.47	2.99	-0.40	**
Number of Siblings	522	2.30	1.64	815	2.39	1.84	-0.09	
PROGRESA	521	0.20	0.40	811	0.15	0.36	0.05	**
B. Kid's Characteristics								
Girl	522	0.50	0.50	815	0.49	0.50	0.01	
Age 2002	522	8.27	1.24	815	8.26	1.19	0.00	
Age 2005	522	11.49	1.19	815	11.39	1.17	0.10	
Std. Cognitive Test Scores	519	0.02	0.94	810	-0.09	1.01	0.12	**
Kid had a disease in 2001	522	0.58	0.49	815	0.60	0.49	-0.02	
Height (cm)	487	126.77	9.80	715	126.88	9.40	-0.11	
Weight (kg)	476	28.02	7.63	694	28.08	7.16	-0.06	
Standarized Height	487	0.08	0.64	715	0.10	0.65	-0.01	
Standarized Weight	476	0.02	0.88	694	0.04	0.92	-0.02	
C. School Characteristics								
Multiple grade school	522	0.20	0.40	815	0.08	0.27	0.11	***
# of teachers at school	522	9.81	4.84	788	10.45	5.44	-0.63	**
# of students registred	512	285.83	174.93	790	329.41	357.67	-43.58	***
Principal's education	521	1.73	0.74	798	1.55	0.67	0.18	***
Principal is a female	522	0.27	0.44	815	0.39	0.49	-0.12	***
Principal has another Job	522	0.33	0.47	815	0.38	0.48	-0.05	*
Cleaning Personnel	522	0.68	0.47	815	0.74	0.44	-0.06	**
Bathroom water/soap wash	500	0.43	0.49	813	0.29	0.45	0.14	***
Ceiling Condition	500	0.86	0.34	813	0.86	0.34	0.00	

[16] Multiple grade schools have students from different grades sharing one classroom and one teacher.

1.3 OLS estimation of the effects of the SBP on children's outcomes

I estimate the effects of the school breakfast program using OLS in both the panel of children from 2002 to 2005 and the cohort of 2002. This allows observing the effect of the SBP in the medium run (4 to 6 years of treatment) and in the short run (1 to 3 years of treatment). The OLS estimation is consistent and unbiased if the selection into treatment is not directly determined either by family's nor children's characteristics influencing children's outcomes, but by other observable and/or unobservable variables not having an influence on the outputs of interest. The intuition behind using OLS is that if the SBP in Mexico is not targeted towards children who are more in need and the assignation is explained by observable school characteristics and other independent unobservables (e.g. political ties to the municipalities), OLS results are potentially consistent. Hence, pooled and simple OLS results are presented as a benchmark before the PSM results are included.

The first set of regressions (i.e. using panel data) include five dependent variables measured in 2005: Raven's scores, a dummy indicating if the child had a disease during the last 12 months, the standardized height and weight by age and sex and a dummy indicating if the child has repeated grade during his primary studies. The breakfast program or treatment status is the main explanatory variable. The regression for the panel takes the form:

$$Yit = \alpha + \beta SBP_i + \gamma_1 C_{it} + \gamma_2 X_{it} + \gamma_2 W_{i2002} + \gamma_3 S_{it} + \pi_i + \mu_{it} \tag{1.1}$$

Where SBP_{it} is the treatment status of the individual i; C_{it} represents a vector of background controls measured in 2002: household's income, parent's education, number of siblings and if the child receives a PROGRESA scholarship or not; X_{it} are the controls for the children' sex and age; W_{i2002} are the controls for the children' outcomes in 2002: Raven's scores, a dummy variable indicating if the child had a disease during the last month and the standardised height and weight of the child i; S_{it} are schools characteristics measured in 2002: a dummy variable showing whether it is a multiple grade school or not, the number of teachers, the number of students, principal's education and sex, and a set of dummies indicating if the school has cleaning personal, bathrooms with water and soap and if school's ceilings are in good condition; π_i are municipality fixed effects and μ_{it} is the error term.

Municipality fixed-effects allow to control for observable and unobservable characteristics at the municipality level that do not change over time and could simultaneously affect children's outcomes and school's treatment status, like the incumbent party in the municipality, infrastructure, quality of institutions, etc. Although the fixed-effects do not address the changes in time across municipalities, the period of time of this study (3 years) is short enough to believe that changes across municipalities do not have big effects on children's outcomes.

The second set of regressions (i.e. using the cohort data from 2002) include the same five dependent variables but measured in 2002. The regression for this cohort takes the form:

$$Yit = \alpha + \beta SBP_i + \gamma_1 C_i + \gamma_2 X_i + \gamma_2 S_i + \pi_i + \mu_i \tag{1.2}$$

The only difference with the Equation (1.1) used for panel data is that it does not include W_{i2002} or the controls for the children characteristics in 2002: Raven's scores, a dummy indicating if the child had a disease during the last 12 months and the standardised height and weight of the child i, since these represent now the outcomes to be measured. Hence only sex and age of children are included in X_i.

Table 1.4 presents the main results of the pooled OLS regression for the five dependent variables. The first column includes the results of SBP with no controls. The coefficients in this column indicate positive and statistically significant effects on Raven's scores and a reduction on the probabilities of having a disease. Column 2 includes context controls and the SBP effects on IQ and health remain, while there is a significant reduction on the probability of repeating grade during primary school. Column 3 includes PROGRESA transfers as a control, and it is possible to observe effects on all the outcomes but height. Nonetheless, in the fourth and fifth columns including a full set of children and school characteristics measured in 2002 as controls, all the effects are non-significant, except for the case of the standardized weight, on which SBP has an effect of 0.11 to 0.13 SD.

Columns 6 to 10 in Table 1.4 show the results of models incorporating municipality fixed-effects. The only significant outcome is the effect on weight. Note that the estimations regarding children's weight are very similar between column 7, which only includes school characteristics as controls, and column 9 which shows the results including a full set of controls (i.e. from 0.17 SD to 0.19 SD). Also, while the estimation in column 8 regarding weight is significantly higher this is only due to the exclusion of the 2002 children characteristics. Consequently, OLS results show that the only consistent effect is that of SBP on weight, while the results for the rest of outcomes are not robust to different specifications and moreover to the inclusion of municipality fixed effects.

Finally, Column 10 presents the effects of SBP including a full set of controls but excluding all children receiving PROGRESA transfers. The results exhibit that for all outcomes, the coefficients do not change dramatically between the full sample presented in column 9 and the subsample in column 10 and moreover the coefficient of weight remains in-line (i.e. approximately 0.16 SD).

Table 1.5 presents the main results of simple OLS regressions using the 2002 cohort data. This regression includes a sample of children observed in the panel but focuses on the 2002 outcomes. The first column shows the regression with no controls. In this column, no statistically significant effects are found. The table also shows the results of the gradual inclusion of context, child's and school's characteristics, as well as school's infrastructure as controls from Columns 2 to 6. It can be noted that the only effect of the SBP is on Raven's cognitive test scores, and this is robust to different model specifications before including fixed-effects. Once municipality fixed-effects are included the effect of SBP on Raven's scores remains significant but only when children receiving PROGRESA transfers are excluded from the sample. However, this significant results should be read carefully since they do not include previous information of cognitive tests as controls.

TABLE 1.4: OLS results of children's outcomes on treatment status and general controls in a panel of students 2002-2005

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Cognitive Test Scores										
SBP	0.147*	0.171**	0.163**	0.117	0.104	0.088**	0.102	0.125	0.044	0.030
	(0.082)	(0.081)	(0.081)	(0.078)	(0.084)	(0.035)	(0.072)	(0.075)	(0.062)	(0.069)
Observations	565	565	565	565	565	565	565	565	565	468
R^2	0.006	0.072	0.074	0.161	0.179	0.150	0.184	0.196	0.259	0.285
B. Illness Probability										
SBP	-0.071*	-0.069*	-0.067*	-0.066	-0.071	-0.068	-0.039	-0.035	-0.043	-0.028
	(0.040)	(0.040)	(0.040)	(0.040)	(0.044)	(0.057)	(0.051)	(0.053)	(0.061)	(0.052)
Observations	587	587	587	587	587	587	587	587	587	487
R^2	0.005	0.017	0.017	0.031	0.049	0.143	0.162	0.176	0.183	0.194
C. Standardized Height										
SBP	0.015	0.068	0.091	0.051	0.089	-0.127	0.002	0.084	0.059	0.056
	(0.083)	(0.081)	(0.079)	(0.062)	(0.068)	(0.081)	(0.100)	(0.091)	(0.096)	(0.104)
Observations	555	555	555	555	555	555	555	555	555	461
R^2	0.000	0.101	0.118	0.487	0.494	0.182	0.263	0.301	0.563	0.558
D. Standardized Weight										
SBP	0.068	0.116	0.126*	0.109*	0.135**	0.124***	0.192**	0.283***	0.174*	0.157*
	(0.079)	(0.077)	(0.076)	(0.056)	(0.063)	(0.044)	(0.077)	(0.070)	(0.091)	(0.079)
Observations	555	555	555	555	555	555	555	555	555	461
R^2	0.001	0.083	0.087	0.517	0.527	0.138	0.190	0.226	0.573	0.560
E. Grade Repetition										
SBP	-0.043	-0.054*	-0.056*	-0.037	-0.040	0.001	-0.017	-0.041	-0.024	-0.025
	(0.031)	(0.030)	(0.031)	(0.030)	(0.033)	(0.061)	(0.054)	(0.050)	(0.059)	(0.060)
Observations	586	586	586	586	586	586	586	586	586	487
R^2	0.003	0.078	0.078	0.141	0.153	0.127	0.153	0.216	0.247	0.275
Context Controls		Yes	Yes	Yes	Yes			Yes	Yes	Yes
PROGRESA			Yes	Yes	Yes			Yes	Yes	
Kid's Controls				Yes	Yes				Yes	Yes
School Controls					Yes		Yes	Yes	Yes	Yes
Municipality F.E.						Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors, clustered on school, in parentheses. Context Controls are parent's education, log of Income and number of siblings; children's controls are age, sex and cognitive test scores, illnesses, height and weight in 2002; School controls are a dummy for multiple-grade schools, number of teachers and students, principal's education and gender and their interactions, if principal has another job or not and school's sanitation/infrastructure. Column 10 excludes children receiving PROGRESA scholarships

TABLE 1.5: OLS results of children's outcomes on treatment status and general controls in a cohort of students in 2002

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Cognitive Test Scores									
SBP	0.107	0.149*	0.149*	0.149*	0.215**	-0.0602	0.102	0.163	0.286*
	(0.0817)	(0.0790)	(0.0797)	(0.0802)	(0.0875)	(0.136)	(0.162)	(0.159)	(0.149)
Observations	587	587	587	587	587	587	587	587	487
R^2	0.003	0.100	0.100	0.100	0.135	0.203	0.232	0.262	0.287
B. Illness Probability									
SBP	0.0302	0.0345	0.0329	0.0326	0.0330	-0.0506	0.0319	0.0302	-0.0202
	(0.0406)	(0.0407)	(0.0409)	(0.0410)	(0.0448)	(0.0646)	(0.0610)	(0.0591)	(0.0611)
Observations	587	587	587	587	587	587	587	587	487
R^2	0.001	0.017	0.017	0.017	0.040	0.157	0.190	0.198	0.231
C. Standardized Height									
SBP	0.0155	0.0680	0.0905	0.197**	0.0891	-0.127	0.00170	0.0591	0.0568
	(0.0836)	(0.0810)	(0.0798)	(0.0851)	(0.0682)	(0.0819)	(0.100)	(0.0961)	(0.104)
Observations	555	555	555	555	555	555	555	555	461
R^2	0.000	0.101	0.118	0.165	0.494	0.182	0.263	0.563	0.558
D. Standardized Weight									
SBP	-0.0491	-0.00603	-0.00450	0.000903	0.0960	-0.0780	0.00241	0.0623	0.175
	(0.0718)	(0.0685)	(0.0682)	(0.0688)	(0.0686)	(0.0919)	(0.119)	(0.116)	(0.121)
Observations	587	587	587	587	587	587	587	587	487
R^2	0.001	0.100	0.101	0.105	0.169	0.211	0.263	0.295	0.321
E. Grade Repetition									
SBP	-0.00236	-0.00943	-0.0101	-0.000170	-0.0201	0.0990*	0.0920*	0.0494	0.0540
	(0.0283)	(0.0278)	(0.0281)	(0.0275)	(0.0299)	(0.0576)	(0.0509)	(0.0416)	(0.0443)
Observations	586	586	586	586	586	586	586	586	486
R^2	0.000	0.053	0.053	0.108	0.123	0.137	0.148	0.264	0.277
Context Controls		Yes	Yes	Yes	Yes			Yes	Yes
PROGRESA			Yes	Yes	Yes			Yes	
Kid's Controls				Yes	Yes			Yes	Yes
School Controls					Yes		Yes	Yes	Yes
Municipality F.E.						Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors, clustered on school, in parentheses. Context Controls are parent's education, log of Income and number of siblings; children's controls are age, sex and cognitive test scores, illnesses, height and weight in 2002; School controls are a dummy for multiple-grade schools, number of teachers and students, principal's education and gender and their interactions, if principal has another job or not and school's sanitation/infrastructure. Column 9 excludes children receiving PROGRESA scholarships.

1.4 Characteristics of the assignment into treatment

This section includes the results of a logistic regression used to analyze the main determinants in the assignation of the SBP to schools. The results obtained here are important to investigate the validity of the Conditional Independence Assumption (CIA) in order to find the causal effects of the SBP through a matching process. The CIA assumes that conditional on X, the selection into treatment and the outcomes are independent; where X is the vector of observed variables included in the logistic model (Heckman, Ichimura, and Todd, 1998). In other words, there is a set of covariates X observable to the researcher and after controlling for these, the potential outcomes are independent of the treatment status or this is "as good as random".

Formally:

$$(Y_i, Y_o) \perp D \mid X$$

The independent variables (X) included in the logistic regression should determine the treatment status (D) and once controlling for this, other unobservable factors should not have influence in the outcomes variables. This is crucial for correctly identifying the impact of the SBP, since it ensures that although treated and untreated groups differ, these differences may be accounted to avoid the selection bias. This allows the construction of a counter-factual for the treatment group, as shown in the next section.

The dependent variable of the logistic regression is a dummy indicating the treatment status; it takes the value of one if the child is in a school participating in the SBP or zero otherwise. The set of independent variables are the same used in the OLS calculations including family context, child's and school's characteristics. I also include a conditional Logit model that allows for the inclusion of fixed effects at the municipality level.

Table 1.6 presents the results of the logistic regressions. The first column including the context variables shows that none of these have a significant influence on the probability of being treated. The second column adds the dummy related to PROGRESA; this is significant and suggests that children receiving this conditional transfers are more likely to be in the SBP. This conclusion does not change once the children's characteristics are included in the third column; however, once school characteristics are included in the fourth column, the PROGRESA dummy is no longer significant. Note that in the specification presented in column 4, the school's characteristics result to be highly significant in determining the probability of being treated. The number of teachers, principals education and other characteristics like the presence of cleaning personnel in the school as well as the hygiene of the bathrooms seem to be the main predictors of the treatment status.

Columns 5 to 7 in Table 1.6 show the results for the conditional Logit regression including municipality fixed effects.[17] The first of these models include only school controls and municipality fixed effects. Despite the fact that most of the variables are non-significant, some school characteristics are still determinant of the assignment into treatment, i.e. the number of students registered and some interactions of principals gender and education. These variables remains significant in the next two specifications when children's (Column 6) and context characteristics (Column 7) are added as controls. Once controlling for context, children and fixed-characteristics of the municipalities, the infrastructure of the school (approximated by the ceilings condition in the classrooms) is also a significant predictor of the treatment.

Notice that despite the local authorities generally declare that the main selection criteria is children's height and weight, such variables are not significant in any of the model specifications. Furthermore, as shown in the Annex Table A1, other Logit models with different measures of height in centimeters, weight in kilograms, Body Mass Index (BMI) and height over weight by age and sex, show that none of these different variables are significant in the assignment into treatment. Similarly, socioeconomic characteristics of the students do not seem to be important determinants of the treatment status.

Consequently, there seems to be evidence of an administrative bias or program placement bias in the application of the SBP based on certain characteristics of the schools. Selection bias may arise from actions on the part of those implementing the intervention (Heinrich, Maffioli, and Vazquez, 2010). Even when there is an explicit targeting criterion, there could be a bias if administrators of the program select those schools more willing to participate in the program, where it is easier to implement it (i.e. in schools with better infrastructure) or where the principal has more empathy with the local authorities.

[17] Given that municipality fixed- effects require of intra-municipality variation across time, children in schools for which there is no variation in some of their characteristics are dropped from the sample, reducing the number of available observations in Columns 5 to 7.

TABLE 1.6: Determinants of the assignment into treatment SBP in Mexico, 2002

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log of Income	0.708	0.669	0.677	0.874			2.965
	(0.484)	(0.483)	(0.512)	(0.734)			(2.304)
Log of Income sq.	-0.044	-0.041	-0.042	-0.053			-0.147
	(0.027)	(0.027)	(0.029)	(0.039)			(0.116)
Mother Schooling	-0.017	-0.015	-0.026	-0.010			-0.104
	(0.036)	(0.036)	(0.037)	(0.042)			(0.090)
Father Schooling	-0.012	-0.006	-0.009	-0.009			0.098
	(0.035)	(0.035)	(0.036)	(0.039)			(0.082)
#Siblings	0.016	0.003	0.019	0.020			0.251
	(0.055)	(0.057)	(0.058)	(0.066)			(0.167)
PROGRESA		0.402*	0.481**	0.096			0.824
		(0.231)	(0.245)	(0.277)			(0.612)
Girl			0.182	0.248		0.481	0.426
			(0.171)	(0.186)		(0.309)	(0.282)
Age			-0.081	0.065		0.420***	0.352*
			(0.075)	(0.082)		(0.156)	(0.184)
IQ			0.165*	0.233**		0.262	0.333
			(0.089)	(0.099)		(0.216)	(0.232)
Standardised Height			0.079	0.166		-0.124	-0.054
			(0.213)	(0.225)		(0.512)	(0.475)
Standardised Weight			-0.047	0.066		0.179	0.234
			(0.144)	(0.152)		(0.292)	(0.290)
Illness			0.136	0.144		0.234	0.446
			(0.174)	(0.194)		(0.294)	(0.332)
Multiple Grade Schools				0.540	0.655	0.648	0.689
				(0.333)	(1.172)	(1.193)	(1.122)
#Teachers				0.267***	0.494	0.534	0.580
				(0.077)	(0.402)	(0.376)	(0.430)
#Teachers Sq.				-0.007***	-0.012	-0.012	-0.012
				(0.002)	(0.016)	(0.015)	(0.017)
#Students				-0.002	-0.011***	-0.013***	-0.014***
				(0.001)	(0.004)	(0.005)	(0.005)
Principal Grad.				0.060	-0.272	-0.151	-0.115
				(0.246)	(1.236)	(1.166)	(1.152)
Principal Posgrad.				1.625***	0.327	0.320	0.459
				(0.370)	(1.188)	(1.064)	(1.086)
Principal is a female				-1.133***	-2.226	-2.517*	-2.419*
				(0.302)	(1.378)	(1.371)	(1.324)
P.Grad*Female				0.990**	1.158	1.298	1.235
				(0.463)	(1.789)	(1.710)	(1.671)
P.Posgrad*Female				-1.635**	-2.442*	-2.396*	-2.704**
				(0.675)	(1.307)	(1.279)	(1.318)
Principal has another job				-0.216	-0.570	-0.546	-0.522
				(0.221)	(0.923)	(0.935)	(0.856)
Cleaning Personnel				-1.044***	0.946	0.763	1.198
				(0.267)	(1.291)	(1.188)	(1.127)
Bathroom Water/soap				0.896***	0.238	0.336	0.078
				(0.221)	(0.534)	(0.558)	(0.544)
Ceilings in good conditions				0.238	1.848	1.975	2.224*
				(0.275)	(1.464)	(1.370)	(1.155)
Observations	587	587	587	587	317	317	317
Pseudo R^2	0.006	0.010	0.019	0.147	0.337	0.375	0.405

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Standard errors, clustered on schools, in parentheses

In regard to this, different authors of the qualitative and quantitative studies conducted in Mexico have mentioned that political affinities with the local authorities are important in the application of the program. For example, Vera (2005) chooses his control group based in four schools that were not treated because the municipality in which they were located was ruled by the leftist party PRD, whereas the state where the study was conducted is governed by the official and incumbent party in Mexico, PRI. The author argues that the reason why the schools did not apply the program was due to political differences with the state authorities.

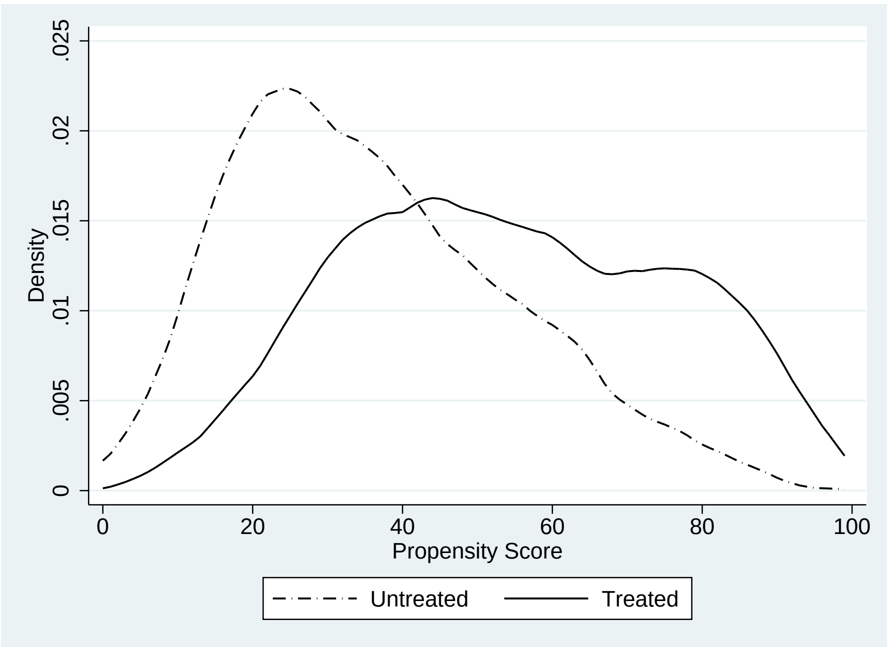
The application of the fixed-effects in the Logit models presented are useful to control for the political characteristics of each municipality, which presumably do not change dramatically in the span of time considered in this study. Once controlling for these time invariant factors, observable characteristics such as principal's education and sex, which arguably reflect the degree of affinity with certain political party, remain significant in the assignment into treatment. Hence, the school characteristics including principal's attributes seem to be a more direct determinant of the treatment status rather than children characteristics and hence, they result to be a main consideration of the CIA and PSM.

1.5 Effects of the School Breakfast Program after Matching

Matching is a widely used method of evaluation. It is based on the idea of contrasting the outcomes of program participants (denoted Y_1) with the outcomes of "comparable" non-participants (denoted Y_0). The differences in the outcomes of the two groups are attributable to the program. The importance of the logistic regression presented above is that it is useful to calculate the propensity score on which the two groups with similar observable characteristics on average can be compared in order to determine the effects of the program. Rosenbaum and Rubin (1983) proved that matching units based on each of the covariates X, is equally valid to match on the propensity score. In other words, the probability of participation in the program summarizes all the relevant information contained in the X variables. In effect, conditioning on the propensity score is equivalent to conditioning on the full vector X, as long as this vector contains all the relevant information to satisfy the CIA (Heinrich et al., 2010).

Figure 1.2 shows the propensity score using the full set of covariates and interactions for both groups, treated and non-treated before matching. Note that a good proportion of the two distributions is already overlapped before the matching procedure. This means that there is a considerable group of children in the sample who share similar family and individual observable characteristics attending similar schools which are treated and untreated, offering a good *common support* for the comparison of outcomes. The probability distribution of the treated represented by the solid line in the graphic, also suggests that given this set of covariates, the program does not seem to be fully directed to one specific group, since the distribution of probabilities of the assignment into treatment are centered around 50 and distributed in a broad range of propensity scores.

FIGURE 1.2: Propensity Score Before Matching



To estimate the treatment effect for each treated person i, outcome Y_{1i} is compared to an average of the outcomes Y_{0i} for matched persons in the control group constructed on the basis of observed characteristics X. Typically, when the observed propensity score of an untreated person is closer to the propensity score of the treated person (i.e. the nearest neighbor) this observations are selected as a match. Alternatively, more than one persons in the control group can also be used as matches (i.e. the five nearest neighbors) and also these matches in the control group can be used for a different treated person (i.e. replacement). Alternatively, the untreated person gets a higher weight in constructing the match if is closer to the treated observation given a specific distance measure. For example, the Kernel methods construct matches using all individuals in the comparing sample and putting less weight in "distant" observations (Heckman et al., 1998, pp.262).

I focus on the measure of the Average Treatment on the Treated (ATT). In principle, given that the SBP should be directed to children in need, the evaluation is not interested in the potential effect of the program on all children, including those more advantaged; nonetheless the ATE results are also presented given the way the SBP was assigned. [18]

Table 1.7 shows the results for the panel of children 2002-2005 after PSM. The outcomes to be measured are the same as in the OLS model presented above, all of them measured in 2005. Column 1 shows the ATT and ATE for a Nearest Neighbour (NN) matching with replacement, using the five closest observations in the control group. Column 2 presents the results for a model using the Mahalanobis distance (MD) of the X covariates.[19] Column 3 shows the results for a radius computation, which establishes that the matched observations from the control group can differ only by 0.001 in their propensity score from those in the treatment group. This could be considered as the least biased estimator of those presented, since the treatment and control groups are very similar in their propensity to be treated. Column 4 shows the results with a slightly wider radius (0.003) which allows increasing the common support between the two groups while keeping balance between treated and untreated units. The last columns present the results of the Kernel non-parametric matching technique which down-weights as matches those units far from the center of the Kernel distribution of each of the treatment units.[20]

The results indicate that the only treatment effect on the treated (ATT) of the SBP is on weight. This result is robust to most of the different specifications presented in Table 1.7. The effect goes from a range of 0.16 standard deviations (S.D.) as presented in column 2 using the MD technique of matching, and up to 0.27 S.D. as presented by what is the least biased estimator in the column 3 (Radius 0.001). Note that all the results presented, except for the MD, suggest that the two groups, treated and untreated, are statistically and jointly equal on average and with 99% of confidence, given the

[18]The Average Treatment Effect (ATE) estimates the mean impact of the program obtained by the average impact across all individuals of the population or, in other words, the effect of the program on an individual of the population randomly chosen

[19]The MD takes into account the correlation in the data, since it is calculated using the inverse of the variance–covariance matrix of the data set of interest. If interest is in the ATT, it uses the variance covariance matrix of X in the full control group. If interest is in the ATE the variance covariance matrix of X considered is for the pooled treatment and full control group

[20]For a broader analysis on different multivariate matching techniques and distance measures see Stuart (2010).

set of covariates used for the matching. Note that despite the MD only achieve the average balance of the treated and untreated groups at the 90% of confidence, the MD estimations do not differ dramatically from the other estimation presented.[21]

Note that OLS results for the panel of children presented in Table 1.4, exhibit the same conclusions than the different specifications of the PSM, supporting a positive effect on weight. The effect of the SBP on weight was also robust to most of the different OLS specifications including the full set of controls and municipality fixed-effects. Finally, in comparison, OLS coefficients regarding the effects on weight go from 0.11 up to 0.28 S.D. These estimations are slightly lower but in line with the PSM results in Table 1.7.

Both, OLS and PSM estimations require that, once controlling for a given set of covariates X, the remaining unobservable variables of the treatment and control groups are not related to the selection into the program nor children's outcomes. If the CIA is accomplished these results offer two important conclusions: on one hand, slightly higher estimations on PSM show that matching treated and untreated groups allows to control for differences between the two groups of children that have an influence on the outcomes which were not considered in OLS. On the other hand, the similarity of the general results with both techniques offer a support to the PSM findings, because arguably, OLS is a technique that involves less decisions made by the researcher in comparison with PSM (i.e. algorithm and bandwidth selection) and in that sense is less vulnerable to a researcher bias.

Figure 1.3 shows the kernel distribution of the propensity scores after matching. The balance achieved by the NN(5) matching seems to be the least effective equalizing both propensity distributions. As expected, the Radius estimation with a caliper of 0.001 as well as the Kernel estimation with a bandwidth of just 0.001 are the most effective matching the two distributions among treated and untreated. In the case of the Radius and Kernel (0.001) the two distributions are skewed to the left, suggesting that the group of treated children who are better matches for the untreated, are those who have a low propensity to be treated (conditional on the set of X covariates) but participate in the program.

Table 1.8 presents the PSM results for the 2002 cohort. These results show the effect of the SBP on children's short-term outcomes, when they have been treated from 1 to 3 years. The six columns of the table include the same matching techniques previously explained and the outcomes to be considered are the same but measured in 2002. The

[21] Details on the matching balance for each of the covariates are presented in the Appendix.

TABLE 1.7: Average treatment on the Treated (ATT) and Average Treatment Effects (ATE) of the SBP on children's outcomes in 2005

	Nearest (5)	Mahala. (5)	Radius (0.001)	Radius (0.005)	Kernel(0.001)	Kernel(0.005)
A. Cognitive Test Scores						
ATT	0.027	0.195*	-0.049	0.076	-0.038	0.055
	(0.160)	(0.108)	(0.117)	(0.055)	(0.117)	(0.111)
ATE	0.053	0.129	-0.056	0.075	-0.049	0.059
	(0.140)	(0.092)	(0.119)	(0.064)	(0.118)	(0.113)
B. Illness Probability						
ATT	-0.154	-0.072	-0.084	-0.083***	-0.096	-0.072
	(0.095)	(0.052)	(0.063)	(0.030)	(0.063)	(0.058)
ATE	-0.057	-0.038	-0.019	-0.048	-0.028	-0.047
	(0.075)	(0.047)	(0.057)	(0.033)	(0.056)	(0.054)
C. Standardized Height						
ATT	0.072	0.103	0.074	0.077	0.072	0.070
	(0.176)	(0.077)	(0.094)	(0.048)	(0.095)	(0.113)
ATE	0.033	0.087	0.036	0.002	0.028	0.006
	(0.136)	(0.071)	(0.084)	(0.052)	(0.085)	(0.100)
D. Standardized Weight						
ATT	0.084	0.162**	0.267***	0.195***	0.257***	0.189*
	(0.148)	(0.080)	(0.091)	(0.045)	(0.091)	(0.114)
ATE	0.094	0.149*	0.162*	0.089*	0.143*	0.095
	(0.131)	(0.081)	(0.087)	(0.051)	(0.086)	(0.103)
E. Grade Repetition						
ATT	-0.084	-0.105***	-0.051	-0.022	-0.038	-0.020
	(0.085)	(0.044)	(0.060)	(0.032)	(0.059)	(0.043)
ATE	-0.053	-0.087**	-0.027	-0.038	-0.021	-0.045
	(0.066)	(0.038)	(0.059)	(0.032)	(0.059)	(0.046)
$p > chi2$	0.758	0.086	0.906	0.463	0.985	0.406
Observations	536	536	536	536	536	536
on-support	505	536	188	392	188	392

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Abadi and Imbens (2006) Robust standard errors in parenthesis.

only significant effects of the SBP are the ATE on Raven's scores in Columns 4 and 6; however, this result does not seem robust to other different matching specifications. The results shown come from the comparison of two groups statistically balanced on average or equal in terms of a full set of covariates. Note that the MD matching, as it considers all the treated and untreated units and not only those in the common support, only achieves the average balance of the two groups with the 90% of confidence.[22]

Jointly, the results presented in Table 1.7 and Table 1.8 suggest that the SBP is probably not having any significant average or treatment effects on any of the analyzed outcomes in the short term. Nonetheless, in the medium run, where the children have already received between 4 and up to 6 years of treatment, the SBP seems to have a positive effect on weight. Finally, OLS results in Table 1.5 focusing on the short term outcomes while controlling for municipality fixed-effects, are closely related to those presented after PSM. In general both, OLS and PSM results offer the same conclusions of no significant effects of the SBP in the short term.

[22] The detail of the matching balance for each of the covariates for this sample of children as well as the graphics of the propensity score after matching for the cohort 2002 can be found in the Appendix.

FIGURE 1.3: Propensity Score After Matching

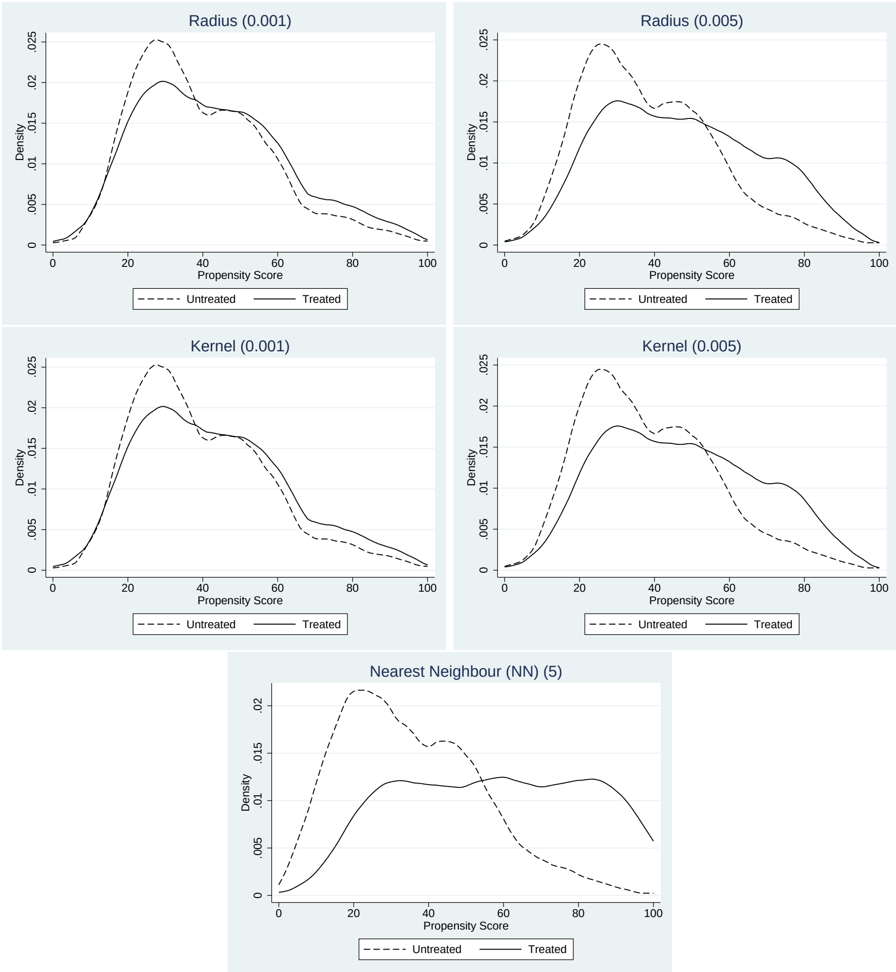


Table 1.9 presents PSM results using the panel data of 2002-2005 but only for the sample of children not receiving a PROGRESA Scholarship. The ATT is less robust to the different specifications of PSM; possibly due to the reduction in the sample size. Nonetheless, the effects on weight remain statistically significant in the MD, the Radius (0.005) and Kernel (0.005). Note that for this sample the MD matching results presented in Column 2 did not achieve the balance of the two groups on average and the point estimator of the ATT (0.13 S.D.) is smaller than the other two significant ATT which present a similar result (i.e. 0.17 and 0.16 S.D.).

For the case of weight, the estimations in the Table 1.9 suggest a smaller effect that in the case of the full sample (where the range was from 0.16 to 0.27 S.D. depending on the PSM algorithm). These results indirectly suggest that the effect of the SBP on weight is higher for children receiving PROGRESA.

TABLE 1.8: Average treatment on the Treated (ATT) and Average Treatment Effects (ATE) of the SBP on children's outcomes in 2002

	Nearest (5)	Mahala. (5)	Radius (0.001)	Radius (0.005)	Kernel(0.001)	Kernel(0.005)
A. Cognitive Test Scores						
ATT	0.092	0.012	0.109	0.093	0.110	0.105
	0.184	0.083	0.120	0.078	0.121	0.080
ATE	0.158	0.052	0.124	0.180***	0.114	0.179***
	0.131	0.080	0.106	0.073	0.107	0.073
B. Illness Probability						
ATT	0.037	0.020	-0.059	-0.023	-0.051	-0.026
	0.079	0.045	0.060	0.042	0.061	0.043
ATE	0.015	0.020	-0.042	0.005	-0.034	0.006
	0.064	0.040	0.057	0.041	0.057	0.042
C. Standardized Height						
ATT	-0.071	-0.028	0.013	-0.046	0.026	-0.042
	0.103	0.045	0.062	0.045	0.062	0.046
ATE	-0.044	-0.017	-0.031	-0.032	-0.026	-0.032
	0.076	0.042	0.059	0.037	0.059	0.038
D. Standardized Weight						
ATT	-0.214*	-0.063	0.030	-0.102	0.042	-0.089
	0.130	0.070	0.096	0.072	0.096	0.072
ATE	-0.118	-0.052	-0.003	-0.062	-0.001	-0.057
	0.104	0.066	0.091	0.062	0.090	0.063
E. Grade Repetition						
ATT	-0.042	0.021	-0.004	-0.028	-0.008	-0.028
	0.067	0.029	0.039	0.029	0.040	0.029
ATE	-0.021	0.022	-0.011	-0.004	-0.008	0.000
	0.049	0.027	0.040	0.029	0.041	0.030
$p > chi2$	0.444	0.027	0.246	0.252	0.165	0.336
obs.	587	587	587	587	587	587
on-support	558	587	234	414	234	414

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Abadi and Imbens (2006) Robust standard errors below point estimates

In order to investigate possible heterogeneous effects of the SBP on children depending on their socioeconomic background, Table 1.10 presents the results of the PSM by the sample of children in the panel 2002-2005 below and above the national median of income in 2005. The results exhibit that for those children at the bottom part of the distribution of household's income, there is a significant effect on weight. By the contrary, for those children in the upper part of the distribution, no robust effects are found on any of the analyzed outcomes. For example, in Column 1, the matching estimation shows a significant ATT of 0.40 S.D. for low-income children, whereas for children at the top half of the distribution of income, as presented in Column 2, the ATT is not significantly different from zero. Columns 3 and 4 show similar results.

Higher positive and significant effects of the SBP on low-income children are also presented in Columns 5 and 7, showing the estimations of the Radius (0.01) and Kernel (0.01) respectively. There, the ATT considering children at the bottom part of the income distribution is of 0.30 and 0.31 S.D with no significant effects on richer children.[23]

[23]It is worth noting that grouping treated and untreated observation by household's income allows to increase the caliper and bandwidth of the PSM to 0.01 (established in the previous PSM estimations to

TABLE 1.9: Average Treatment on the Treated (ATT) and Average Treatment Effects (ATE) of the SBP on Kid's outcomes in 2005 in a sample of kids without PROGRESA transfers

	Nearest (5)	Mahala. (5)	Radius (0.001)	Radius (0.005)	Kernel(0.001)	Kernel(0.005)
A. Cognitive Test Scores						
ATT	-0.011	-0.071	-0.052	0.082	-0.055	0.097
	0.180	0.103	0.131	0.078	0.133	0.079
ATE	-0.096	-0.084	-0.113	0.017	-0.120	0.014
	0.146	0.090	0.131	0.080	0.133	0.083
B. Illness Probability						
ATT	-0.129	-0.055	-0.029	-0.067	-0.029	-0.069
	0.089	0.052	0.075	0.041	0.076	0.043
ATE	-0.068	-0.042	-0.045	-0.052	-0.043	-0.057
	0.071	0.046	0.071	0.046	0.072	0.047
C. Standardized Height						
ATT	0.011	0.043	0.009	-0.048	0.040	-0.062
	0.179	0.077	0.110	0.069	0.110	0.070
ATE	0.021	0.013	0.006	-0.043	0.035	-0.047
	0.131	0.074	0.105	0.070	0.106	0.071
D. Standardized Weight						
ATT	0.130	0.131*	0.082	0.171***	0.121	0.163***
	0.158	0.076	0.116	0.062	0.115	0.064
ATE	0.121	0.109	0.095	0.119*	0.117	0.120*
	0.132	0.081	0.110	0.064	0.109	0.065
E. Grade Repetition						
ATT	-0.052	-0.050	-0.023	-0.072**	-0.024	-0.063**
	0.072	0.039	0.053	0.030	0.053	0.030
ATE	-0.041	-0.057*	-0.016	-0.061*	-0.021	-0.057
	0.057	0.032	0.054	0.034	0.055	0.036
$p > chi2$	0.051	0.001	0.407	0.582	0.359	0.411
obs.	556	556	556	556	556	556
on support	540	556	211	418	211	418

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Abadi and Imbens (2006) Robust S.E. below point estimates

In conclusion, the effect on weight for lower income children is higher than that of the full sample of children. The significant results shown in Table 1.7 estimate an ATT from 0.16 to 0.26 S.D. Whereas in the estimation for the low income children, the effect ranges from 0.30 to 0.40 S.D. These results offer evidence that the SBP may have stronger effects on the weight of children in the lower distribution of income.

1.6 Conclusions

The SBP has been at the core of the alimentary strategy of the Mexican government for many decades; but the evaluation of its results has been scarce or limited, even more so at the national level. This research focused on evaluating the impact effects on outcomes such as cognitive, school and anthropometric measures. OLS and PSM estimations with

0.001 and 0.005) without including "bad" matches as controls. This also allows to keep a good sample size for the PSM estimations

TABLE 1.10: Average Treatment on the Treated (ATT) and Average Treatment Effects (ATE) of the SBP on Kid's outcomes in 2005 in a sample of kids below and above the median of income

	Nearest (5)		Mahalanobis (5)		Radius (0.01)		Radius (0.01)	
	Below	Above	Below	Above	Below	Above	Below	Above
A. Cognitive Test Scores								
ATT	0.029	0.059	0.381**	0.117	0.061	0.090	0.074	0.078
	0.205	0.294	0.172	0.138	0.155	0.227	0.158	0.226
ATE	0.030	0.016	0.182	0.137	0.118	-0.064	0.127	-0.077
	0.184	0.215	0.141	0.128		0.179		0.177
B. Illness Probability								
ATT	-0.155	-0.268*	-0.039	-0.169	-0.074	-0.130	-0.080	-0.132
	0.110	0.150	0.066	0.106	0.075	0.119	0.077	0.120
ATE	-0.037	-0.161	-0.043	-0.055	0.004	-0.138	0.007	-0.138
	0.091	0.125	0.061	0.079		0.097		0.097
C. Standardized Height								
ATT	0.314	-0.092	0.391***	-0.079	0.220	-0.207	0.239	-0.217*
	0.199	0.233	0.106	0.120	0.161	0.131	0.165	0.129
ATE	0.254	-0.084	0.183*	0.022	0.184	-0.228	0.204	-0.230
	0.172	0.192	0.096	0.108		0.140		0.140
D. Standardized Weight								
ATT	0.398**	0.063	0.331***	0.058	0.297**	-0.093	0.314**	-0.123
	0.187	0.227	0.101	0.135	0.141	0.144	0.144	0.145
ATE	0.363**	0.034	0.238**	0.085	0.281	-0.095	0.287	-0.099
	0.167	0.220	0.107	0.116		0.145		0.146
E. Grade Repetition								
ATT	-0.087	-0.044	-0.145**	-0.055	-0.058	-0.023	-0.061	-0.027
	0.100	0.111	0.063	0.082	0.067	0.062	0.068	0.064
ATE	-0.060	-0.065	-0.116**	-0.071	-0.027	-0.074	-0.029	-0.082
	0.081	0.075	0.054	0.050		0.055		0.055
$p > chi2$	0.867	0.973	0.020	0.026	0.355	0.712	0.314	0.757
obs.	324	212	324	212	324	212	324	212
on support	286	173	324	212	234	121	234	121

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Abadi and Imbens (2006) Robust S.E. below point estimates

different algorithms were presented. The results suggest that the assignation into treatment does not depend on children's characteristics like height and weight, moreover, they suggest an administrative bias in its assignment. The estimations also show a consistent effect on children's weight which may push them above their standardize average by age and sex.

Specifically, results coming from Logit models do not support that selection into treatment depends on children's observable characteristics like their height and weight despite local authorities' presumption of such assignment criteria. By the contrary, school characteristics seem to be more related to the application of the SBP suggesting the presence of an administrative bias. As discussed in Section 1.4, variables like school's infrastructure and the closeness of the principals with the local authorities might explain part of this bias. This is perhaps not surprising considering that the SBP in Mexico is an old program born and designed in the surge of the so-called *corporative State* in Mexico, commonly related to the use of social programs for political purposes.

The program does not seem to present any significant effects on height in any of the different estimations in the OLS and the PSM with different logarithms, neither in the short (1 to 3 years of treatment) nor the medium run (4 to 6 years of treatment). This is unsurprising since the available evidence on children's growth show that faltering in developing countries typically begins in the first three months of life and recovery (catch-up growth) after early childhood is difficult to observe (Eveleth and Tanner, 1976; Grantham-McGregor, Cheung, Cueto, Glewwe, Richter, and Strupp, 2007; Victora, de Onis, Hallal, Blössner, and Shrimpton, 2010). Although recent literature suggests that puberty may offer an opportunity window for recovery (Coly, Milet, Diallo, Ndiaye, Bénéfice, Simondon, Wade, and Simondon, 2006; Prentice, Ward, Goldberg, Jarjou, Moore, Fulford, and Prentice, 2013; Hirvonen, 2013), children in the SBP are neither in early childhood nor puberty (i.e. 6 to 12 years old children).

In theory, a better nutritional status in the short run can have impacts in the long run health and this might be reflected in lower children's dropout rates and less absenteeism. Some researchers have found immediate effects of breakfast on children's nutritional status, including a positive effect on child's size (see Powell et al. (1998); Cueto and Chinen (2008)). However, such evaluations do not include any specific measure of disease probabilities. The results reported here do not show any effect of the SBP on the probabilities of having a disease after 1 to 3 or 4 to 6 years of treatment. In general, the sign of the coefficients obtained in the OLS and PSM show a reduction on the probabilities of having a disease, but the standard errors of these estimations do not allow to conclude any significant effect.

Grade repetition probabilities do not seem to be significantly affected by the presence of the SBP in neither the short nor the long-run. Although the sign of the coefficients is always negative in the PSM, suggesting that the probability of repeating a grade reduces for the children in the program, the results are generally not robust to the different algorithms presented. Regarding the effects of the SBP on grade repetition, several studies have found a positive relation between school feeding programs and enrollment or attendance rates in low-income countries in the short term and specifically in undernourished children (Del Rosso, 1996; CARE, 2004; Grantham-McGregor, 2005; Cueto and Chinen, 2008). However, as mentioned in Section 1.1, the percentage of undernourished children in Mexico is only 1.6%, which can explain the absence of a significant effect of the SBP on grade repetition.

In general, PSM results do not show any positive results on children's cognitive tests. OLS estimations support a positive effect in the short run; however this is not confirmed by the PSM estimations in neither the short nor the medium run. These results are not surprising since recent evidence on children's cognitive skills show that gaps in children's abilities that play an important role on diverse outcomes, open up very early across socioeconomic groups (i.e. before the age of 5) and these differences remain into adulthood. Evidence shows that interventions in early stages of life are the most important closing IQ gaps; however gains on Raven's scores fade out around 4 years after the intervention, and only in earlier more intensive interventions (1 to 2 years old) cognitive gains last until early adulthood (Cunha, Heckman, Lochner, and Masterov, 2006; Heckman, 2008; Heckman, Pinto, and Savelyev, 2012).

Finally, PSM estimations show a consistent effect of the SBP on weight after 4 to 6 years of treatment. The results presented exhibit an ATT of 0.16 to 0.26 S.D. for the full sample of children. Nonetheless, it is worth noting that treated children in this sample are already 0.02 S.D above their average by age and sex as shown in Table 1.3. This means that after 4 to 6 years of treatment these children could be pushed further above their standardized average weight as a result of the SBP. More specifically, these effects could be even more detrimental on children at the bottom half of the income distribution. In this regard, a qualitative study on a sample of treated students from the State of Mexico (the most populated state in the country) suggested, after reviewing children's consumption habits, that most of SBP children do have breakfast at home, later in the day, they also receive breakfast at school as part of the program and moreover, some of them buy even more food at school (see Shamah et al., 2010). This evidence gives support to SBP effect found in this research of a higher chance of overweight in children.

Chapter 2

Does lengthening the school day increase academic achievement? Evidence from a natural experiment

Economic research shows that improvements on math, language and science test scores relate to increases in economic growth, earnings in adulthood and to the reduction of the inequality of income between social groups (Murnane et al., 1995, 2000; Lazear, 2003; Hanushek and Kimko, 2000; Barro and Lee, 2001; Hanushek, 2004, 2013) Additionally, other non-monetary benefits from education such as improved health status and lowered crime have also been reported.[1] Nonetheless, the 2009 results of the Programme for International Student Assessment (PISA) shows that, Mexico is located in the 48th place in reading and 50th in math out of 65 countries members and partners of the Organization for Economic Co-operation and Development (OECD).

Similarly, in the case of the Test for the National Assessment of Academic Achievement in Schools (ENLACE, for its abbreviation in Spanish), which evaluates math and language skills of all Mexican children in basic education, the results are not very promising either. In 2009, around 70% of Mexican students in primary education exhibited results which are considered 'insufficient' or 'elementary' in both subjects. Undoubtedly, this implies a significant and challenging problem for educators to ensure that future generations do not suffer from the severe basic skills problems that currently hinder many children. However general evidence on education remains unable to offer a

[1] For a recent review of the available evidence on this matter see Lochner (2011)

clear guidance of what policies and specific investments should be pursued to increase educational outcomes, even less so in specific contexts such as the Mexican.

Several policies directed to schools have shown to increase enrollment, however, experimental and non-experimental research regarding diverse public interventions, has not provided strong evidence on pupil's achievement. Fee reductions, conditional transfers and school nutrition programs in developed countries have exhibited effects in enrollment which alas, are not accompanied by increased achievement. Other policies related to overall expenditures and school initiatives such as lower class size and more educated teachers are not conclusive in their relation to students outcomes (Hanushek, 2003).[2] Similarly, the positive impacts on learning reported in developing countries come from few variables such as availability of desks, teacher's knowledge and teacher absence, which provide little guidance for future policy and programs (see Glewwe, Hanushek, Humpage, and Ravina, 2011).

In response to the weak evidence about the impact of an increased educational spending, governments have turned their attention to policies that modify the way schools are run and organized. For example, by decentralizing schools' decisions to the level of local governments and schools rather than national or state bureaucrats[3] or by increasing the lenght of the school day along with a modification in the structure of teaching.

The idea that increasing instructional time is expected to promote learning and achievement via increased time on task, broader and deeper coverage of curriculum, more opportunities for experimental learning and deepened adult-child relationships, is a central notion in education that has been broadly discussed in the United States (US) (Link and Mulligan, 1986; Levin and Tsang, 1987; Brown and Saks, 1987; Slattery, 1995; NECTL, 2005) and more recently in Latin America (see Holland, Alfaro, and Evans, 2015). Some examples of this type of programs are the *No Child Left Behind* act in the US that stimulates the allocation of extra time to teaching math and reading; the *Future for Education and Care* in Germany that provides funding for full-time schools; the *Extended School Times* project in the Netherlands and the *Full-time School Programs* recently implemented in Latin American countries such as Chile and Uruguay.

The current study focuses on the impact analysis of a program of increased hours applied in basic schools of Mexico known as the Full-time Schools Program (*Programa*

[2]Although, evidence from experimental evaluations have found some evidence of a positive effect from a reduction in class size (Angrist and Lavy, 1999)

[3]In this regard, a few studies offer evidence of positive effects on test scores and school attendance of school decentralization programs in Argentina, Mexico, Bolivia and Colombia (Galiani, Gertler, and Schargrodsky, 2008; Skoufias and Shapiro, 2006; Faguet and Sánchez, 2008)

Excuelas de Tiempo Completo, or PETC) on primary academic achievement,[4] measured in standardized test scores of mathematics and Spanish from the 2008-2009 to the 2012-2013 academic year.[5] PETC seeks to improve learning opportunities by increasing the time children spend at school from four and a half to eight hours everyday, while incorporating new subjects and activities in the curricula (e.g foreign languages, arts, culture and nutrition) and granting every year a fixed stipend for operative expenses and a varying fund according to the number of professors and students in each school. Every primary school may participate in the program, but PETC is supposed to target disadvantaged and rural schools.

The program started in the 2007-2008 academic year in 500 basic schools located in 15 out of Mexico's 32 States.[6] By 2013, 6715 basic schools from all the country were participating in the program (i.e. approximately 10% of all basic schools that can potentially be included). This represents a spending of about US$460 million from 2007 to 2013. Moreover, the 2012 elected federal government has announced an expansion of the program from 2013-2014 in order to reach 40,000 primary and secondary schools by 2018. According to the Secretariat of Finance in Mexico, the budget programmed for 2014-2015 rose US$1 billion. Nevertheless, to the best of my knowledge, the most recent expansion of PETC has been dictated without any previous public evaluation of the potential causal impacts on school and children's outcomes such as test scores and grade repetition at the national level of this large program that is aimed to be an important component of the educational strategy in Mexico.

The present research combines different sources of information to generate a novel and large census dataset including the database of the ENLACE test, PETC administrative data and school-level information coming from a yearly census survey conducted in basic schools (better known as *statistics 911*). These statistics include a wide range of characteristics such as number of students, professors' and principals' level of education as well as instructional time in Arts, IT, and foreign languages, along with information on family expenses required by schools on educational materials.

[4]The study excludes secondary education despite being also affected by the program because grades 9th to 12th are taught in a broad range of institutions, such as Technical Secondary Schools, State Secondary Schools, Federal Secondary Schools, and "Telesecundarias". Each of them already use different time schedules ranging from 5 hours in "Telesecundarias" to 6-8 hours in Technical Secondary Schools. All of these institutions can participate of PETC, therefore, the effect of the program on time extension is different. Even though this variation results interesting to analyze, with the data at hand, it is not possible to identify the different time schedules applied in each secondary school.

[5]From now on academic years are denoted also as years, so for example, 2008 refers to 2007-2008 academic year.

[6]By 32 States, I refer to Mexico's 31 federal entities and the Federal District located in Mexico City.

A parallel evaluation to PETC conducted by Andrade-Baena (2014) uses DiD and PSM separately, to evaluate the impact on ENLACE test scores using administrative information and characteristics of the municipalities where schools are located. The author finds positive effects ranging from 0.06 SD to 0.13 SD and 0.07 SD to 0.13 SD for Spanish and mathematics, respectively. Nonetheless, the study reports significant differences between controls and treatment groups before PETC introduction (i.e. 'placebo tests') and these remain significant after including controls. This threat could be the result of the definition of the control group along with the quality of the regressors included (i.e. at the municipality level and only for 2010). The present research differentiates from Andrade-Baena (2014) by the inclusion of school level information and the further analysis on the impact channels of PETC.

The methodology applied in this research takes advantages of the gradual application of the program in the period from 2009 to 2013 as a natural experiment and uses DiD to arguably obtain causal effects on achievement separated by years of treatment (i.e. one and up to four years of treatment). Two reasons define the period to be analyzed: a) ENLACE test scores are fully accountable and comparable from 2008-2009 onwards;[7] and b) schools from the first cohort treated by PETC (2007-2008) included units that already had different versions of extended times of instruction (e.g. 'Escuelas de Jornada Extendida') and these schools could have been working as such from one up to ten years before PETC introduction; furthermore, these schools are not clearly identified.

The identification strategy relies on the fact that selection into the program is independent of the trends on the average outcomes that treated and control groups exhibit before and after the program started. In other words, although average test results and grade repetition are different between PETC and control schools, both groups show a parallel trend in outcomes before policy intervention. Furthermore, in order to avoid further concerns of unobserved heterogeneity not captured in the DiD models due to the higher variation observed in the characteristics of the schools used in our large control group included in the main estimation (57000 schools, approximately), the strategy is refined by the computation of a PSM that pairs similar schools between the original treated and control groups. The objective is to reduce the chance of including schools that are very different to those treated (e.g. much richer/larger or poorer/smaller), hence new DiD estimations are obtained with a more homogeneous control group.

The contributions of this study are threefold. First of all, it contributes to the scarce empirical literature on the estimation of causal impacts of extended hours in schools.

[7] Specific characteristics of this test will be discussed in detail in Section 2.2

Secondly, it differentiates from previous works by using census data and test scores from all primary schools in a country and not from a sample. Thirdly, this study is the first to offer evidence of the effects of PETC on the average and for different subgroups (i.e. with high marginality) and can be used as a reference to evaluate future extensions and targeting of the program in Mexico and for its implementation in other developing countries.

Estimates show average effects close to 0.06 SD on mathematics and 0.07 SD on Spanish test scores. Results also show a significant and positive effect on the standardized test scores of both subjects, ranging from approximately 0.04 SD after two years of treatment to 0.11 SD after four years of treatment on math and from 0.05 SD to 0.11 SD, respectively, on math scores of a panel of schools with a full set of school characteristics as controls. These effects are robust to different specifications, the application of ’placebo tests’, examination of different treatment and control groups and the matching of control schools with similar observable characteristics. Further inspections on causal channels show that PETC has a higher impact after four years of treatment (0.29 SD) on both subjects in schools with high marginality and exhibits a positive effect on children at the botton and at the top of the scores distribution. Results also show that the program does not have an effect on dropout rates nor in the selection of ”better” students, arguably suggesting that the effects do not come from changes in the composition of students in treated schools.

The rest of this chapter is presented as follows. Section 2.1 discusses prior evidence on full-time school programs. Section 2.2 outlines the main characteristics of PETC since its inception. Section 3.2 presents the data and includes descriptive statistics. Section 2.4 discusses the empirical strategy and presents the main results. Section 2.5 discusses some of the impact channels of PETC on test scores. Section 3.6 concludes.

2.1 Prior evidence

Prior evidence on the extension of the school day remains scarce and shows, at worst, no effect on test scores and at best, a small relationship between instructional time and student academic achievement. Research suggests that the relationship is stronger for students with initially low academic achievement while displaying diminishing effects of increasing instructional time on student test scores (Wheeler, 1987; Bishop, Worner, and Weber, 1988; Adelman et al., 1996). Findings also suggest that as the measure of time is refined to more closely reflect the amount of time devoted to the outcome analyzed, the relationship was strengthened (Caldwell, Huitt, and Graeber, 1982), and that only time spent successfully completing instructional activities and not allocated time, has a relationship with achievement (Levin and Tsang, 1987; Karweit, 1985). Hence, this policy could be more effective when considerations are made for how time is used, including classroom management, the appropriateness of instruction and curriculum, and student motivation (Aronson, Zimmerman, and Carlos, 1999) [8]

Nevertheless, there are many methodological limitations in most of the previous studies. Longitudinal and rigorous research on time in school is lacking, and existing studies have been repeatedly challenged for being weakly designed, based on correlational data and case studies (Cuban, 2008). Several studies make use of small and non-randomly selected samples and are based on cross-sectional data. Moreover, although some studies have examined the same classrooms or schools at different times, most of them have considered relatively short periods of time, typically less than an academic year (Bellei, 2009). Finally, it is not clear to what extent these studies controlled for confounding factors that may bias the estimates. As a consequence, the literature revealed that designs are generally weak for making causal inferences (Patall et al., 2010).

A handful of studies arguably allow for causal inference indicating neutral to small effects. For example, Robin (2005) estimates the impact of preschoolers attending an extended time program in a urban district of New Jersey. A total of 294 low-income students were randomly assigned to pre-school programs of different durations. Children either attended the experimental program in a public school for 8-hours per day, 45 weeks per year or during half-day, 3.5 hours and 41 weeks. Students in the experimental program outperformed children in the control group in both math and literacy.

James-Burdumy, Dynarski, Moore, Deke, Mansfield, Pistorino, and Warner (2005) evaluate the 21st Century after-school centers, a program that typically offered homework sessions, academic activities, enrichment activities, such as art, drama, or music, and recreation activities across 12 school districts in the US. The authors randomly assigned students either to a treated (1,258 students) or to a control group (1,050 students). The intent-to-treat (ITT) impacts, as well as the local average treatment effect (LATE) show that neither the effects on teacher assigned grades in math and English, nor standardized reading test scores were significant. Although, subgroup estimates of

[8] A detailed review of the prior evidence on day extension and number of days spent in school per year can be found in Patall, Cooper, and Allen (2010)

ITT impacts suggest a positive effect on English grades for students with low initial reading test scores.

Meyer and Van Klaveren (2013) conduct a randomized field experiment to estimate the effect of an extended day program in seven Dutch elementary schools included in the Extended School Times project on math and reading achievement. Empirical results of this study show no significant effect on either of the two measured outcomes.

For the case of developing countries, Bellei (2009) takes advantage of the gradual implementation of the Chilean full-time schools program and uses it as a natural experiment to calculate Difference-in-Difference (DiD) estimators and evaluate the impact on the academic achievement of high school students. The results exhibit a small but positive and significant overall effect on language tests of 0.05 to 0.07 standard deviations (SD) and a no effect on math in a period of two years. The evidence also suggests that the program had larger positive effects on rural students, students who attended public schools and students located in the upper part of the achievement distribution.

Likewise, Cerdan-Infantes and Vermeersch (2007) estimate the impact of the full-time school program in Uruguay on standardized test scores of 6th grade students. The program was not randomly placed but targeted to poor urban schools, hence, authors use propensity score matching (PSM) to cope with the selection problem and construct a comparable control group. The results show that students in disadvantaged schools improved their test scores by 0.07 SD per year of participation in the full-time program in math and 0.04 SD in language.

2.2 PETC Characteristics, Selection and Testing patterns

PETC started in the 2007-2008 academic year aiming to improve learning opportunities, diet and ensuring retention of children in basic education by extending the school day from four and a half to eight hours in all public schools of basic education. As a consequence, this policy increases instructional time to 1200 class-hours distributed in 200 days per scholar year. From its inception, PETC aimed to increase not only the amount of instructional time dedicated to core subjects such as reading and math, but it also included six work lines aiming to achieve a holistic education and to develop lifelong competences: a) fostering learning of curricula contents; b) didactic use of information and communication technologies (IT); c) learning of additional languages; d) art and culture; e) healthy life; and f) recreation and physical development (UNESCO, 2010; Gómez, Flores, and Alemán, 2013). This way, the program seeks to give teachers more time to consolidate reading, writing, oral expression, critical thinking, scientific and mathematical thinking with the use of IT and teaching of a second language. The program also seeks to improve children's feeding and studying habits with the inclusion of a cafeteria, meals and specific time to help them develop better learning and study skills (SEP, 2010).[9]

Although the curricula for PETC schools is flexible, the program allows for a specific time (i.e. one hour at the end of the school day) for teachers to plan and evaluate their activities and, if necessary, talk to parents. The program guidelines for schools also suggest specific hours everyday to tutor students and help them with their homework during the eight hours at school.[10]

For the purposes of the program, schools should preferably have a dining room, a computer classroom and sports infrastructure. This has represented a total spending for the federation of approximately US$460 millions from 2007 to 2013 invested on reconditioning schools with computer classrooms, roofed patios, laboratories, kitchens, dining halls and toilets. This budget also covers the training and monetary aids for principals, teachers, and support staff members; monitoring, didactic materials, meal's services and supplies (Gómez et al., 2013). Unfortunately, there is no public data available on the costs per school for all the years used in this study but it was possible to obtain from the budget office in SEP, an approximate amount of money granted to an average school is of approximately US$40,000 of which around US$15,000 are fixed. [11]

A comparable program in terms of fixed costs is the Quality Schools Program (Programa Escuelas de Calidad PEC, for its abbreviation in Spanish) that offers grants of US$15000 per school and aims to decentralize educational decisions at the school level and to foster communities participation. PEC has shown some results on attainment (that is, grade repetition and dropout rates) but results on performance are uncertain Skoufias and Shapiro (2006). As it will be explained below, PEC is of great relevance for our analysis as Full-time schools should be part of the Quality Schools Program.

[9]Secondary objectives of the program include to allow working mothers to extend their workday, to support mono-parental families and to prevent at-risk students from engaging in harmful activities such as drugs and crime (SEP, 2010, p.3)

[10]An of example of the timetable suggested for PETC schools can be found in Table B1in the Appendix.

[11]In general terms, the formula used multiplies US$290 per moth per teacher, US$350 per month per principal, and close to US$25 per month per student.

Possible threats to the objectives of the program are covered by a study of characterization conducted by UNESCO (2010), which surveys 953 principals in full-time schools in 2008. Some key results are that 81% of the activities covered during the extended time are conducted mainly by the same teachers who were hired pre-intervention, while the rest of activities are taught by new external specialists and teachers. This may well imply an extra load of work for teachers that could compromise their quality. Additionally, only 60% of the schools report to have received a visit by the technical board at least once a year and a low 40% declare to have received specific training for the implementation of the program. Finally, given that it is not mandatory for students to stay the eight hours at school, 10% of them do not stay during the full school day. Regardless, 90% of principals consider that the program favors the implementation of new pedagogical strategies and improves students learning, 86% believe that student's satisfaction has improved, 76% that students applications increased and 75% consider that PETC should be mandatory in all basic schools in Mexico, because it helps students to enhace their competences and it also allows to put more emphasis on students and other pedagogical activities.

2.2.1 How were PETC schools selected?

Schools selected into PETC from 2008 to 2012 should have generally completed a list of requirements based on (SEP, 2010), these include:

- Schools should be participating in the Quality Schools Program ("Programa Escuelas de Calidad" PEC, for its abbreviation in Spanish). PEC is a program seeking to decentralize educational decisions to the school level rather than the federal or state level, giving more participation to the general community. This program is directed to rural, indigenous and urban schools with high levels of marginality. PEC schools are planned to be in the program from 1 to 5 years depending of the needs of each school. This is a key factor in the consideration of the treatment and control groups as discussed in the next section.[12]
- There exists a Technical Board in the State where the school's are located, which will supervise and follow the implementation of the program.
- The community is open to participate in the activities of the full-time schools (e.g. offering support in the dining rooms).

[12] For more details on PEC, see Skoufias and Shapiro (2006).

- Schools have minimum infrastructure requirements (e.g. space for the construction of kitchen and computer classrooms, sports infrastructure, and basic services such as water and electricity).
- Schools are only working in one shift either in the morning or afternoon but not both. In Mexico, approximately 40% of primary schools offer two shifts. This is also considered further in the construction of the control group.
- Preferentially, schools should be located in vulnerable geographic areas.

Once eligible schools have been identified by the federal authorities, according to the aforementioned requirements, potential schools to be treated are suggested to each of the 32 States. Nonetheless, it is worth mentioning that the list of potential schools only work as a guideline and each State can lastly define the schools included in PETC. Once they have been selected, the implementation of PETC consists of two stages: 1) the organization and preparation of schools previous to their inclusion to the program (i.e. infrastructure, teachers and staff hiring) and 2) the design, organization and development of the teaching objectives. In the second stage, teachers receive printed materials which suggest pedagogical strategies to be implemented during the extra-time at schools and to develop the competences necessaries for the instruction of new contents. Along with it, State's Technical Boards evaluate and support the implementation of the program with 'regular' visits to the primary schools (UNESCO, 2010).

The program started in the academic cycle 2007-2008 in 500 primary and secondary schools located in 15 out of the 32 federal entities; by 2009, 953 schools were treated in 29 states; 2,000 schools were participating in 2010; 2,273 in 2011; 4,758 in 2012 and by 2013, 6,715 were included in all Mexico.[13] These numbers represent more than 10% of the approximately 62,500 schools which can potentially be included in PETC, according to the requirements referred above (CONEVAL, 2013).

2.2.2 ENLACE Test and patterns of application

ENLACE is a census standardized exam of mathematics and Spanish (plus one extra subject, i.e. science or history rotating every year) directed to evaluate knowledge and

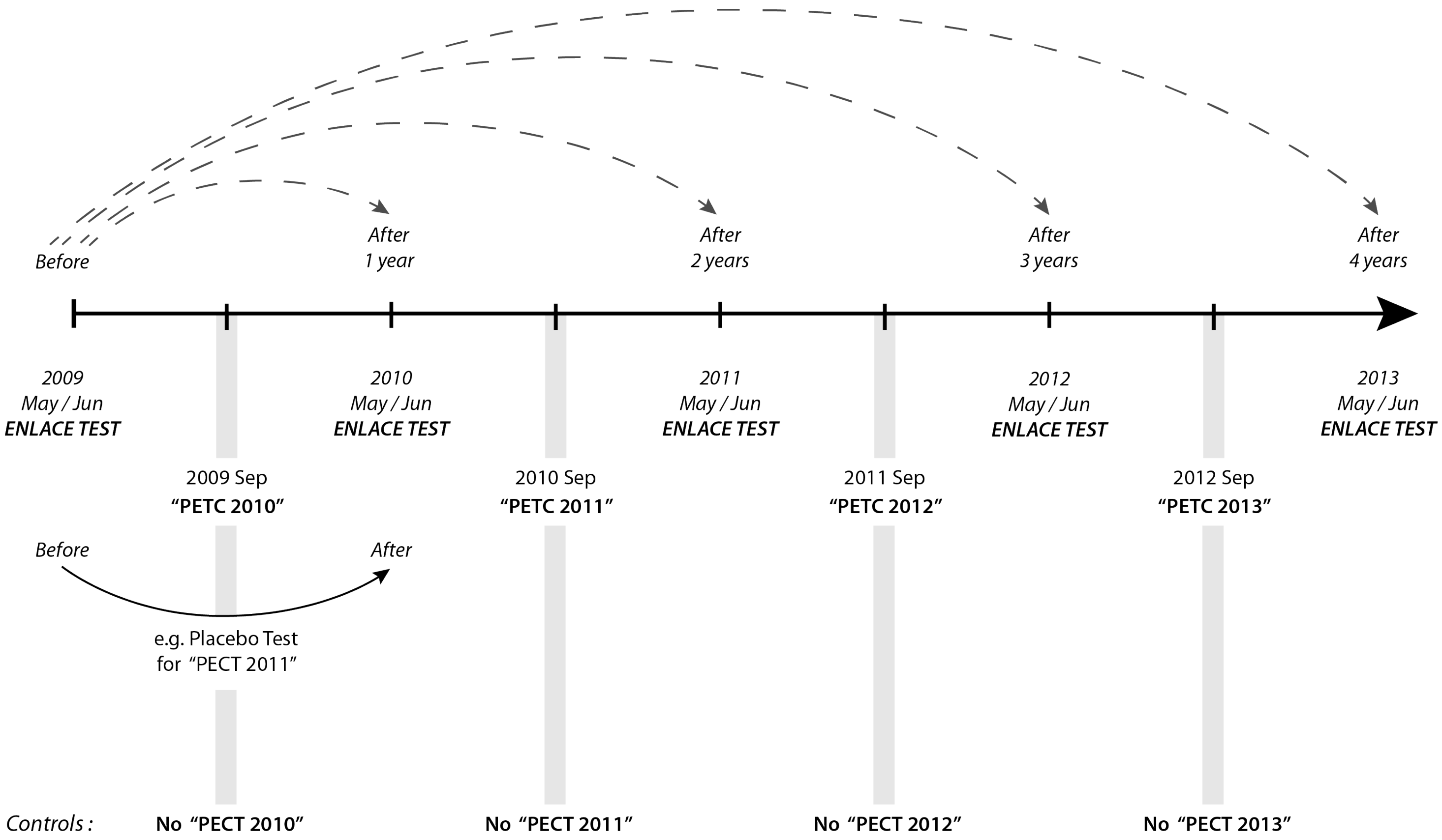
[13]Note that these numbers are based on treated, pre-scholar, primary and secondary schools, but since this study will only focus on primary schools, the final number of treated schools will be lower as shown in the descriptive statistics presented in Section 3.2.

skills of students from third to sixth grade of primary education, first to third grade of secondary education and first year of high school. The results of the test are expressed in a standardized scale comparable through time (200 to 800 points with an average of 500). ENLACE has been applied to both public and private schools since the academic cycle 2006-2007. Nonetheless, the test was fully accountable and comparable between years only after 2008-2009, when the staff conducting the test started to be completely unrelated to the school where ENLACE was taking place.

The test is applied every year in a short period of time either in the last week of May or during the first week of June. Handily, the PETC schools start their scholar year in September and finish in July. Given this configuration of time, it is possible to observe test results before and after schools have entered the program in more than one period of time.

As shown in Figure 2.1, the data at hand allows us to observe ENLACE results for the scholar cycle 2008-2009 (test applied in May/June 2009) of the schools that will enter the program in September of 2009 (named as PETC 2010). These schools are tested again in May/June 2010, after one scholar year of treatment, and subsequently until May/June 2013, after 4 years of treatment. This pattern of application allows the construction of different control and treatment groups and placebo tests, since ENLACE results are available before and after PETC schools started the program in 2011, 2012 and 2013.

FIGURE 2.1: Timeline: Pattern application of ENLACE and PETC



2.2.3 Definition of control and treatment groups

Between 2010 and 2013, treated schools are defined as those entering the program in each specific year, whereas the controls are defined as schools which can *potentially* be treated but have never been treated and remain untreated during the whole period here analyzed. Potentially treated schools are defined for the purpose of this research, as general public primary schools operating only in one shift. Alternatively, a second control group is built from the original controls. The basic method used is that of Heckman et al. (1998), where propensity scores are estimated for the ten nearest neighbors with no replacement and common support, and the sample is then trimmed to exclude poorly matched schools. School's observable characteristics are useful to perform this exercise. Propensity score is an attempt to further standardize the set of treatment and control schools.[14]

As mentioned before, PETC schools are required to be in PEC and can be participating in the latter as much as one and up to five years, depending on the time each school require to fully decentralize its operations. Table 2.1 shows the total number of schools participating in both programs. Effectively, contrary to what is stated by the PETC requirements, not all the schools that belong to PETC belong to PEC. For example, in 2010, 290 schools or 37% of the treated by the full-time schools program do not belong to PEC. For this reason, two variables are defined to identify schools in both programs: one identifies the total number of years the schools have been in PEC by the moment they start participating in the full-time schools program (this variable act as a control in the regressions I will define in the next section). A second variable identifies schools that have been at least one year in PEC during the analyzed period, this works to identify heterogeneous effects of PETC in schools with and without PEC.

Table 2.1 also identifies the number of schools treated, controls and the matched controls to be included in this study. For example, the potential group of schools analyzed for 2010 is formed by 776 treated; 53,044 control schools and 5,137 matched schools integrating the second control group, however, during the course of this research all estimations will be presented for the pooled treatment and control groups.

[14]The probit models including the variables used for PSM as well as balancing tests for each cohort of PETC schools can be found in Tables B7 to B14 in the Appendix.

TABLE 2.1: Treated and Control Primary Schools Participating in Schools Quality (PEC) and Full-Time Schools Program (PETC) 2010 to 2013

				Schools used as control group					
	Treated			All non-PETC			With Matching*		
	PEC	*No PEC*	*Total*	*PEC*	*No PEC*	*Total*	*PEC*	*No PEC*	*Total*
(2009-2010)	365	411	776	25,188	27,856	53,044	3,231	1,906	5,137
	47%	*53%*	*100%*	*47%*	*53%*	*100%*	*63%*	*37%*	*100%*
(2010-2011)	143	122	265	25,471	28,005	53,476	1,255	658	1,913
	54%	*46%*	*100%*	*48%*	*52%*	*100%*	*66%*	*34%*	*100%*
(2011-2012)	1,135	793	1,928	24,351	27,639	51,990	6,768	3,193	9,961
	59%	*41%*	*100%*	*47%*	*53%*	*100%*	*68%*	*32%*	*100%*
(2012-2013)	327	189	516	25,239	27,953	53,192	1,958	594	2,552
	63%	*37%*	*100%*	*47%*	*53%*	*100%*	*77%*	*23%*	*100%*

Source: author's elaboration based on PEC and PETC administrative data.
* Probit regressions are used to predict the linear index of the propensity score for the sample of PETC schools and all non- PETC schools. Units within the 'common support' are selected for difference-in-differences analysis.

2.3 Data and Descriptive Statistics

The empirical analysis is based on a novel dataset that includes different sources of information: a) the results of ENLACE test; b) school census data (known as *statistics 911*); and c) the administrative data of PETC and PEC which identify the schools treated in both programs. All data sources combine at the school level for the period 2009 to 2013. As discussed, this rich dataset allows to observe an important number of schools' characteristics relevant to the analysis conducted.

The results of ENLACE for each of the schools and students are published by SEP. This dataset include the average results by subject, the percentage of students with levels of insufficient, fair, good and excellent, as well as the number of students tested and unreliable tests per school.[15] The geographical location of the schools: state, municipality and locality is also reported along with five categories of 'privation' or marginality suffered in school's localities.[16]

The *statistics 911* are self reported questionnaires sent by the schools to SEP at the beginning of each scholar year. They include information on number of students by grade,

[15]Every year a set of questions to be used in the next year's test is applied to a controlled sample, this works to built the standardized scale of the next year's test and allows to identify students out of this scale who are labeled as unreliable. Furthermore, ENLACE includes quality controls through an automatic validation to detect collusion with the use of the models K-Index and Scrutiny as described in technical details of the ENLACE manual.

[16]The level of marginality is calculated by the National Council of Population (CONAPO, for its abbreviation in Spanish) and it is based in eight socioeconomic variables of the locality where the school is located, considering: average education levels, household's characteristics (i.e. available services and infrastructure) and goods availability. For further details see CONAPO (2010)

age and sex, number of students who passed and failed, number of classrooms, information of basic services such as water and electricity, number of teachers, administrative personal and teachers' and principals' level of education. These data can be combined with ENLACE in order to have information about school's performance.

A third source of data is the administrative databases of both PEC and PETC, which serve to identify treated schools, shift, region, municipality and locality where these are located. Both administrative data sources are also provided by SEP.

2.3.1 Descriptive statistics

Table 2.2 shows the main descriptive statistics of the pooled sample of treated and untreated schools from 2008 to 2013. Panel A shows information of variables related to the ENLACE test. Note that treated schools have a significantly higher number of students tested. Proportionally, the number of students tested and with unreliable results is significantly lower in treated schools (at the 10% level of significance).

Panel B shows that, on average, treated schools have participated almost twice as many years in PEC than untreated schools and this difference is highly significant. In general, treated schools have more students, teachers, administrative workers and more classrooms. More importantly the marginality index is relatively lower in treated schools (2.36) than in control schools (2.75), suggesting a better socioeconomic context for students in treated schools. On average, there are more principals with postgraduate education present in treated schools (0.21 vs. 0.14 in control schools). Also, note that the proportion of teachers with bachelors and postgraduate education is higher in PETC schools.

Panel C show the instructional time of 'non-core' activities in schools: sports, artistic education, IT and English as second language. Unfortunately, time dedicated to core subjects such as mathematics, reading and science is not reported. The statistics show that on average, treated schools spend more time on these subjects, specially on the teaching of a second language and sports. Panel D includes figures showing average family spending. Differences in spending on books and fees are not statistically different between treated and control institutions, this is not surprising since all primary schools are publicly funded. However, average spending in uniforms (usually not provided by the State) is slightly higher in treated schools (35 pesos, or approximately US$2.5 per year).

In general, these numbers suggest that treated schools are different from the controls in observable and unobservable ways. PETC schools are bigger and feature a slightly higher proportion of teachers with a professional career and postgraduate studies. On average, treated schools also seem to be located in a better socioeconomic environment.

A circumstance that may well explain why PETC schools seem to be in a better position is that SEP can only suggest the potential schools to be treated but each State can choose the schools that the local government believe are more suitable for the treatment. It is possible then, that the States are choosing those schools which are easier to access (e.g. those closer to the municipality offices) or those which already have the infrastructure to run the program. These units may well be located in geographic areas with a better socioeconomic environment. This is something that is taken into consideration in the methodology to evaluate the impact of the program, controlling for school characteristics including their marginality index and by the computation of a propensity score based on the observable characteristics of schools, as it will be discussed in detail in section. 2.4.

TABLE 2.2: Main descriptive statistics by treatment status from the pooled sample: 2008 to 2013

...

	All non-PETC Schools					*PETC Schools*						
	Mean	S.D.	Min	Max	N	Mean	S.D.	Min	Max	N	Difference	
A. ENLACE Test												
# students tested	109.04	112.35	1.0	1210.0	336961	130.91	94.72	1.0	752.0	11278	21.87	***
# tests untrusted	6.00	11.64	0.0	551.0	336961	6.07	10.46	0.0	180.0	11278	0.07	
% students tested	93.26	34.95	0.3	100.0	301998	93.87	8.40	5.2	100.0	10150	0.61	*
% untrusted tests	3.36	7.28	0.0	100.0	301998	3.17	6.07	0.0	75.0	10150	0.19	*
B. School characteristics												
PEC (years)	1.28	1.93	0.0	6.0	530226	2.63	2.32	0.0	6.0	12360	1.35	***
Students	179.69	177.10	1.0	2531.0	415842	221.29	148.65	3.0	1146.0	10968	41.60	***
Principals	1.00	0.31	0.0	3.0	417589	1.02	0.30	0.0	3.0	10926	0.02	***
Teachers	7.08	5.19	1.0	30.0	364107	8.18	4.71	1.0	30.0	10625	1.10	***
Administrative workers	1.10	1.67	0.0	15.0	417563	2.53	3.09	0.0	15.0	10762	1.43	***
# classrooms	6.88	4.40	0.0	17.0	385598	8.39	4.13	0.0	17.0	10016	1.51	***
Marginality Index	2.75	1.41	1.0	5.0	336961	2.36	1.35	1.0	5.0	11278	-0.39	***
# of Principals by education												
Vocational	0.38	0.51	0.0	3.0	417779	0.35	0.50	0.0	3.0	10928	-0.03	***
Bachelors	0.46	0.52	0.0	3.0	417694	0.45	0.52	0.0	3.0	10929	-0.01	*
Postgraduate	0.14	0.35	0.0	3.0	417836	0.21	0.42	0.0	2.0	10932	0.07	***
% of Teacher's by education												
Vocational	36.89	34.02	0.0	100.0	363474	34.01	30.13	0.0	100.0	10563	-2.88	***
Bachelors	56.89	34.66	0.0	100.0	363476	59.66	30.40	0.0	100.0	10563	2.77	***
Postgraduate	5.31	13.46	0.0	100.0	363474	5.67	12.85	0.0	100.0	10563	0.36	**
C. Instruction Time (h/week)												
Sports	3.61	6.11	0.0	20.0	365605	5.60	6.87	0.0	20.0	8268	1.99	***
Artistic education	0.61	2.71	0.0	20.0	414151	1.27	3.96	0.0	20.0	10666	0.66	***
IT education	0.47	2.51	0.0	20.0	414315	1.09	3.92	0.0	20.0	10614	0.62	***
Second language	0.63	2.90	0.0	20.0	405463	2.33	5.41	0.0	20.0	9731	1.70	***
D. Spending (pesos/year)												
Books	285.94	915.03	0.0	80000.0	418660	290.29	1348.92	0.0	70000.0	11011	4.35	
Uniforms	362.67	1106.67	0.0	99800.0	418634	397.93	1742.78	0.0	90000.0	11012	35.26	**
Fees	203.28	1035.29	0.0	98000.0	418443	208.70	931.31	0.0	50750.0	11002	5.42	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3 provides descriptive statistics for the outcome variables of interest: the standardized test scores of Spanish and math before and after the application of PETC for each of the treatment and the two control groups from 2010 to 2013: all primary schools which can potentially be treated and a smaller control group including the ten nearest neighbors of each treated school according to a PSM.

Test measures are higher on average in PETC schools at the base time and after treatment. For example in 2010, considering the pre-policy year, treated schools where 0.240 SD above the average in math results, while the controls are 0.184 SD above. Once a set of matched controls is constructed, differences become smaller and the outcomes appear to be more similar for the comparison groups of 2010, 2011 and 2012. Although, in the case of the matched controls in 2013, differences seem to remain considerable.

For valid inference to be drawn, it is necessary to show that baseline differences in the pre-policy period have remained stable in years previous to the policy intervention (to ensure a“like with like” comparison). Further evidence on the parallel trends of outcomes before PETC is presented in the empirical approach contained in the next section. Bearing this in mind, DiD results presented in Table 2.3 should be read carefully, but the figures suggest a recurrent non-significant difference between the outcomes of treated and controls before and after PETC (one year of treatment). More importantly, size and significance does not vary considerably when the comparison is made to the matched controls.

TABLE 2.3: Mean outcomes for various samples

	Number of Schools	Standardized Test Scores Pre-policy	Post-policy	Change	DiD	
Mathematics PETC 2010						
Treated	721	0.240	0.257	0.017		
All non-PETC schools as controls	49808	0.184	0.206	0.022	-0.006	(0.022)
Matched controls	4928	0.260	0.282	0.022	-0.006	(0.022)
Spanish PETC 2010						
Treated	721	0.217	0.253	0.036		
All non-PETC schools as controls	49808	0.178	0.195	0.017	0.020	(0.022)
Matched controls	4928	0.260	0.275	0.015	0.022	(0.024)
Mathematics PETC 2011						
Treated	219	0.376	0.423	0.047		
All non-PETC schools as controls	51135	0.207	0.219	0.012	0.036	(0.043)
Matched controls	1875	0.342	0.361	0.019	0.028	(0.045)
Spanish PETC 2011						
Treated	219	0.335	0.387	0.052		
All non-PETC schools as controls	51135	0.196	0.208	0.012	0.041	(0.041)
Matched controls	1875	0.332	0.352	0.02	0.033	(0.049)
Mathematics PETC 2012						
Treated	1883	0.348	0.17	-0.178		
All non-PETC schools as controls	49885	0.214	0.069	-0.145	-0.032	(0.017)
Matched controls	9872	0.372	0.240	-0.132	-0.029	(0.018)
Spanish PETC 2012						
Treated	1883	0.364	0.181	-0.183		
All non-PETC schools as controls	49885	0.202	0.044	-0.158	-0.025	(0.016)
Matched controls	9872	0.383	0.219	-0.164	-0.02	(0.017)
Mathematics PETC 2013						
Treated	490	0.399	0.416	0.017		
All non-PETC schools as controls	47111	0.071	0.106	0.035	-0.019	(0.031)
Matched controls	2495	0.196	0.263	0.067	-0.051	(0.031)
Spanish PETC 2013						
Treated	490	0.431	0.469	0.038		
All non-PETC schools as controls	47111	0.047	0.086	0.039	-0.002	(0.031)
Matched controls	2495	0.247	0.328	0.081	-0.043	(0.034)

For all non-PETC schools as controls, standard errors are clustered on school; for matched controls, these are clustered on school and bootstrapped with 100 repetitions and no replacement

2.4 Impact of PETC on Test Scores

This section evaluates the impact of PETC on test scores and grade repetition using DiD models. This method is based on the Wald estimator and has been broadly described and used in a number of earlier papers.[17] DiD seeks to control for a large number of observable factors and for unobserved school heterogeneity. Considering these factors is important, owing to the different levels of pre-policy achievement in test scores and grade repetition between PETC and control schools as discussed. In effect, different observed and unobserved factors such as the socioeconomic context, marginality of schools and infrastructure, can explain the difference in results before and after policy intervention.

[17] See for example Heckman and Robb Jr (1985); Machin and McNally (2008); Hussain (2012)

Additionally, it is also important to consider that changes after policy intervention are related to PETC rather than to the historic trends observed in the outcomes. Hence, the basic estimates are derived from the following model:

$$Y_{st} = \beta PETC_s + \gamma t_s + \delta_1 (PETC_s * t_s) + \delta_2 X_{st} + \pi_e + \mu_{st} \tag{2.1}$$

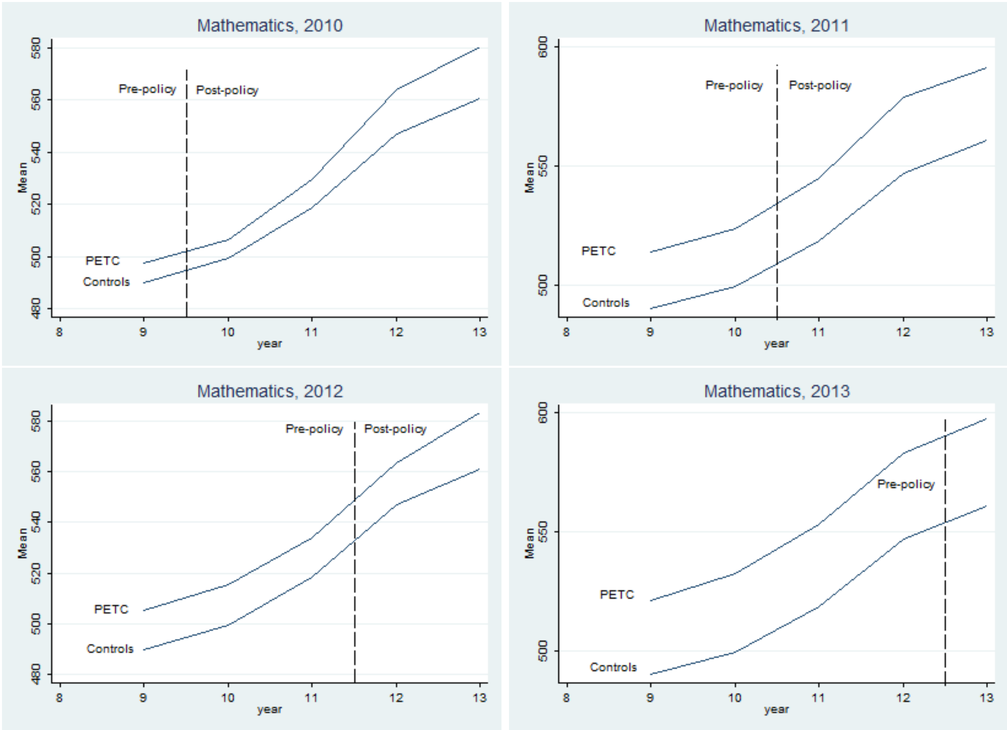
Where Y_{st} is the outcome of interest for school s in time t; β_s accounts for the differences between treatment and control group (PETC is a dummy equal to one for schools in the program); γ is a time trend common to control and treatment groups. PETC is interacted with t_s which is set equal to one for the time period when the PETC policy was in effect and zero in pre-policy period. The coefficient δ_1 is the DiD estimate of the PETC policy; δ_2 captures the influence of a vector of controls X which includes characteristics of schools such as the number of students and classrooms and a marginality index, instruction time in arts, sports, IT and languages, principals' and teachers' education and family's spending on schools materials, along with variables indicating the proportion of students taking the ENLACE test by school and the proportion of results considered as 'unreliable', as well as the years schools have participated in PEC; π_e denotes regional fixed-effects and μ_{st} is an error term.

Since school differences in the pre-policy period are included in the model captured in β_s, what is measured are within-school changes in test outcomes and grade repetition before and after PETC introduction in treatment schools relative to within-school changes in the outcomes of control schools. However, the critical requirement to achieve an unbiased DiD estimator is the *parallel-trend* assumption. Formally, the error term: $cov(\mu_{st}, PETC * t_s) = 0$, or in other words, the changes in the outcome of interest between treated and untreated units should not be explained by other factors previous to the introduction of the policy (i.e. outcomes could have already been increasing faster for treated schools previous to PETC).

Figure 2.2 shows the raw average trend of math results. Treated schools have higher scores in all periods and roughly, the trends for the four treatment and control groups appear to share the same tendency before the application of PETC. For the first treatment group (2010), the graphic is useful to observe the post-policy trends, suggesting a small positive change for PETC schools. The graphic of the last treated and control groups

(2013) is more useful to review trends previous to policy intervention, which appear to be parallel.[18]

FIGURE 2.2: Trends of ENLACE mathematics average scores by treatment status



[18]Similar results are observed in the graphs for the average results of Spanish and the matched control groups of math and Spanish. These can be found in Figures B1, B2 and B3 in the Appendix.

2.4.1 Basic DiD results

Table 2.4 shows the average effects of PETC on mathematics test scores for treated schools compared to non-PETC schools and a matched control group. The first column presents the "raw" effect of a DiD model without any controls, on average and by separating the effects in years since policy intervention. Results show that treated schools present a significant difference respect to non-treated of 0.038 SD. First column also shows a pattern of increasing impacts through time ranging from a non-significant effect during the first year of treatment and up to 0.78 SD after four years of treatment.

Column 2 shows the effects of a DiD with a full set of school characteristics as controls. The average effect of the policy is higher compared to column 1, indicating that the characterisitcs of schools do interact with policy effectiveness. Similarly, during the first year since policy intervention, there are no effects on math test scores. Nonetheless from the second year of treatment PETC schools show a positive effect on average ranging from of 0.036 SD growing to 0.111 SD four years after policy intervention.

Column 3 displays the results for the matched non-PETC schools according to the observable characteristics of schools. Results do not differ dramatically and keep the same pattern observed in column 2, on average and by years of treatment, becoming stronger after two (0.046 SD) and up to four years of treatment (0.107 SD).

TABLE 2.4: Basic Results: PETC on Mathematics Standardized Test Scores

	(1)	(2)	(3)
Control Schools	All non-PETC schools		With Matching
PETC * Policy On	0.038**	0.059***	0.061***
	(0.015)	(0.014)	(0.013)
*PETC * 1 year after policy*	0.014	0.017	0.025*
	(0.013)	(0.012)	(0.015)
*PETC * 2 years after policy*	0.020	0.036**	0.046***
	(0.016)	(0.015)	(0.015)
*PETC * 3 years after policy*	0.043*	0.066***	0.060**
	(0.023)	(0.021)	(0.025)
*PETC * 4 years after policy*	0.078***	0.111***	0.107***
	(0.023)	(0.023)	(0.027)
Control variables	No	Yes	Yes
School fixed-effects	Yes	Yes	Yes
Number of Schools	164,520	164,520	59,569
R^2	0.003	0.164	0.146

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Columns 1 and 2 show standard errors, clustered by school, in parentheses. Column 3 shows bootstrap standard errors from 100 replications, 100% of replacement and clustered on school, in parentheses.

Table 2.5 shows the results for the effects of PETC on Spanish. It presents the same specifications than Table 2.4. The raw effects (column 1) show a significant average effect for any treated school of 0.054 SD, higher than what was observed for mathematics. No significant effects are found in column 1 after one year of intervention but similarly to the results on mathematics test score, from the second year of treatment there is a significant and cumulative effect of the policy ranging from 0.033 SD to 0.108 SD four years after policy intervention.

Column 2 shows significantly higher effects on average (0.073 SD) and by years after policy intervention, being small but significant from the first year of treatment (0.021 SD) and up to 0.137 S.D. after four years. Note that this results are rather similar when comparing PETC schools to statistically matched non-PETC schools on average and by years of treatment, as presented in column 3.

TABLE 2.5: Basic Results: PETC on Spanish Standardized Test Scores

	(1)	(2)	(3)
Control Schools	All non-PETC schools		With Matching
PETC * Policy On	0.054***	0.073***	0.067***
	(0.015)	(0.013)	(0.014)
*PETC * 1 year after policy*	0.018	0.021*	0.027**
	(0.013)	(0.012)	(0.013)
*PETC * 2 years after policy*	0.033**	0.049***	0.050***
	(0.015)	(0.014)	(0.016)
*PETC * 3 years after policy*	0.059***	0.080***	0.069***
	(0.022)	(0.020)	(0.022)
*PETC * 4 years after policy*	0.108***	0.137***	0.111***
	(0.022)	(0.022)	(0.024)
Control variables	No	Yes	Yes
School fixed-effects	Yes	Yes	Yes
Number of Schools	164,520	164,520	59,569
R^2	0.004	0.181	0.160

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Columns 1 and 2 show standard errors, clustered by school, in parentheses. Column 3 shows bootstrap standard errors from 100 replication, 100% of replacement and clustered on school, in parentheses.

2.4.2 Placebo Tests

This section presents placebo tests that allow to discard significant differences in outcomes between treated and untreated schools before PETC which could be explained by unobserved factors, once a set of controls is included.

Table 2.6 shows the results of placebo regressions for all outcomes. Columns 1 to 3 show the results for all non-PETC schools used as controls, while columns 4 to 6 show the coefficients for the controls after PSM. Columns in the table show the results of the diferent treatment cohorts, while the rows show the effects up to three years before they were treated. This way, column 1 shows the DiD coefficient for math and Spanish between PETC schools treated in 2011 and their counter-factual one year before they were treated. Hence, the data allow to observe DiD results between treated and untreated units up to two years before, in the case of schools that started the program in 2012, and up to three years before for the PETC schools treated in 2013.

Results in columns 1 to 3 for mathematics, show that there are no significant differences between treated and control schools in the pre-policy period. Note that, once schools are matched, PETC schools in 2013 appear to have a significant difference in math results compared to their controls three years before they were treated (0.067 SD in 2009-2010); however, this difference disappears for the coming years.

Regarding Spanish results, similar conclusions can be drawn for PETC schools starting in 2013. For both type of regressions, including all the controls and only matched controls, there is a significance difference three years before policy introduction (2009-2010 in columns 3 and 6). In both cases, this significant difference happens three years before the program started and disappears for the coming two years before PETC 2013. In general, the results in Table 2.6 only suggest a possible threat for the conclusions of the effects on Spanish test scores, specifically for schools treated in 2013.

TABLE 2.6: Placebo regressions: DiD and PSM-DiD for all outcomes

	Math Scores					
	DiD			*PSM and DiD*		
	(1) PETC 2011	(2) PETC 2012	(3) PETC 2013	(4) PETC 2011	(5) PETC 2012	(6) PETC 2013
One Year Before Policy	-0.058	-0.026	-0.042	-0.054	-0.014	-0.015
	(0.037)	(0.022)	(0.030)	(0.041)	(0.011)	(0.028)
Number of Schools	98641	98659	95520	3674	22130	5411
Two Years Before Policy		0.007	-0.003		0.020	0.026
		(0.012)	(0.027)		(0.013)	(0.025)
Number of Schools		98659	99463		22155	5417
Three Years Before Policy			0.040			0.067***
			(0.028)			(0.025)
Number of Schools			98637			5434
	Spanish Scores					
	DiD			*PSM and DiD*		
	(1) PETC 2011	(2) PETC 2012	(3) PETC 2013	(4) PETC 2011	(5) PETC 2012	(6) PETC 2013
One Year Before Policy	-0.025	-0.005	-0.001	-0.018	-0.003	0.022
	(0.035)	(0.011)	(0.030)	(0.043)	(0.010)	(0.029)
Number of Schools	98641	99463	95511	3674	22130	5411
Two Years Before Policy		0.004	-0.009		0.013	0.001
		(0.011)	(0.029)		(0.011)	(0.024)
Number of Schools		98659	99440		22155	5417
Three Years Before Policy			0.058**			0.072***
			(0.029)			(0.027)
Number of Schools			98637			5434

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
DiD regressions show standard errors, clustered on school, in parenthesis. PSM and DiD regressions show robust standard errors from 100 replications, 100% of replacement and clustered on school, in parentheses. Regressions include a full set of controls, including school's teachers' and principals' characteristics as well as controls for the number of years in PEC, marginality of the school area and dummies for six mexican regions.

2.4.3 Heterogenous effects of PETC

In general PETC seems to have a positive effect on test scores, however it is still important to consider possible heterogeneous effects of PETC. It is plausible to think that the average positive effect of the policy may well be explained by “the best” schools doing better without having much effect on more deprived schools which may well on average have less motivated and/or skilled students and account with less resources to make the extra time of teaching effective. This could be judged as a negative result if it translates into an increase in the gap between relatively poorer and richer schools. Furthermore, it is important to consider the fact that some schools are presenting different effects depending on their participation in one or two of the substantially important educational programs in Mexico, PETC and the Schools Quality Program (PEC), as discussed above.

Table 2.7 present heterogeneous effects by schools marginality and PEC participation. Columns 1 and 2 show the average effect of PETC on mathematics and Spanish test scores compared to all non-PETC schools separated by their level of marginality.[19] The results exhibit a positive a significant effect for both type of schools and on both subjects, but it is clearly stronger for more deprived schools or with a higher index of marginality. For example, PETC schools do 0.166 SD better in mathematics and 0.162 SD in Spanish compared to non-PETC schools with high marginality. This contrasts to lower gains of 0.037 SD and 0.049 SD, respectively, in low marginality PETC schools.

Finally, columns 3 and 4 show slightly higher average effects for schools participating in both programs, moreover in the case of mathematics when PEC plus PETC schools present gains of 0.046 SD after policy intervention compared to non-significant effects on schools only participating of PETC.[20]

TABLE 2.7: Heterogenous Effects: PETC on Mathematics and Spanish Standardized Test Scores by level of marginality and PEC participation

	(1)	(2)	(3)	(4)
	Low Marginality	High Marginality	Only PETC	PEC plus PETC
A. Mathematics				
PETC * Policy On	0.037**	0.166***	0.034	0.046***
	(0.015)	(0.033)	(0.026)	(0.017)
Number of Schools	90,586	73,939	82,518	82,007
R^2	0.161	0.158	0.170	0.153
B. Spanish				
PETC * Policy On	0.049***	0.162***	0.047*	0.063***
	(0.014)	(0.032)	(0.025)	(0.016)
Number of Schools	90,585	73,935	82,513	82,007
R^2	0.180	0.142	0.180	0.173
Control variables	Yes	Yes	Yes	Yes
School fixed-effects	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Standard errors, clustered on school, in parentheses.

Results of the effects of PETC on test scores separated by low and high marginality schools and by years since policy intervention are plotted in Figure 2.3. It can be observed that altough the effect on low marginality schools grows over time, this remains lower than the improvement presented in more deprived schools. In effect, while low marginality PETC schools exhibit a positive and significant effect of 0.05 SD in mathematics and 0.07 SD in Spanish four years after intervention, more deprived schools present a significantly higher average gain of 0.29 SD in both subjects.

[19]Note that the proportion of treated schools with low marginality is 70% while a considerable 30% of treated schools belong to more deprived localities.

[20]All PETC effects on math test scores separated by cohort and years of treatment can be found in the Appendix Table B2 using all non-PETC schools and in Table B3 using a matched control group. For the case of Spanish these can be found in the Appendix Table B4 and Table B5, respectively.

Considering that the average math scores of treated schools in the pre-policy period is 513 points with a SD of 63 in low marginality schools and 463 with and SD of 80 in high marginality schools (a difference of 50 points) these effects translate into a marginal gain of only 3.2 points for more advantaged schools, while it represents a gain of 25 points for deprived schools, that is almost half of the pre-policy gap between high and low marginality schools. For the case of Spanish, with an average of 456 for high marginality schools (SD of 70) and 510 for low marginality schools (SD of 57), the gains for deprived schools translate into aproximately a third of the gap between more advantaged and disadvantaged institutions before policy introduction.

FIGURE 2.3: Average effects of PETC on mathematics and Spanish standardized test-scores by school's marginality and years of treatment

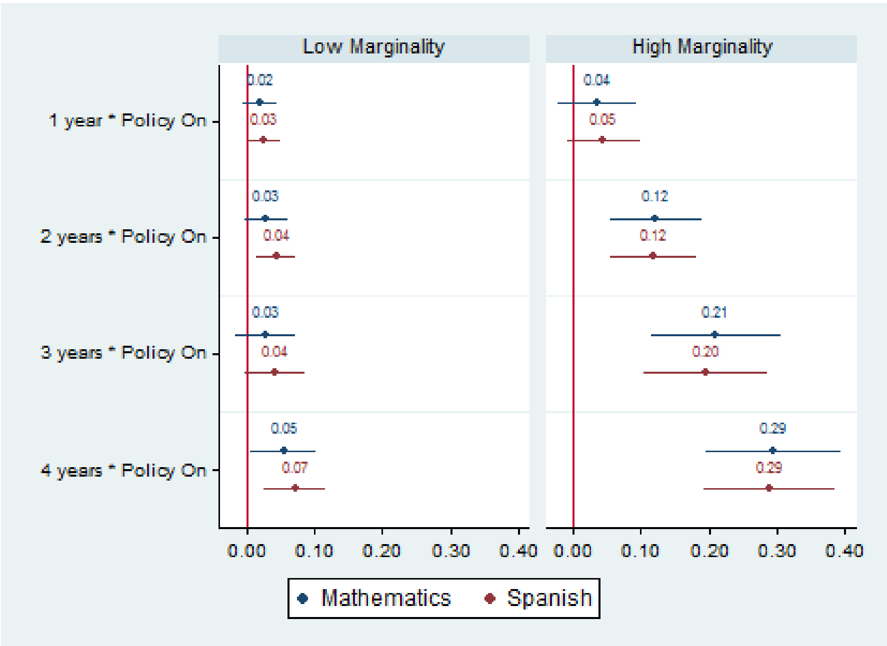


Figure obtained from the point estimators and the 95% confidence intervals coming from a DID regression including a set of dummy variables interacting a post policy dummy with the number of years since intervention. The regression also includes school fixed effects, time-fixed effects and full set of school characteristics as controls. The counter-factual is constructed from all non-PETC schools.

Figure 2.4 shows heterogeneous effects by PEC status. The results show a different pattern suggesting that after 2 years of treatment PEC plus PETC schools have a higher impact on test scores but this difference reduces and practically disappears after three and four years post-policy. Furthermore the effects on mathematics are lower for schools participating of both programs (0.07 SD) compared to PETC schools (0.10 SD). Hence in the medium-run, joint effects of PEC and PETC are not additive and participating only in the full-time schools program seems as effective for school's improvement as the participation in both programs. As mentioned, both programs are similar in their fixed costs, however, PEC does not seem to be adding more to what PETC alone is doing in terms of gains in test scores.

FIGURE 2.4: Average effects of PETC on mathematics and Spanish standardized test-scores by PEC participation and years of treatment

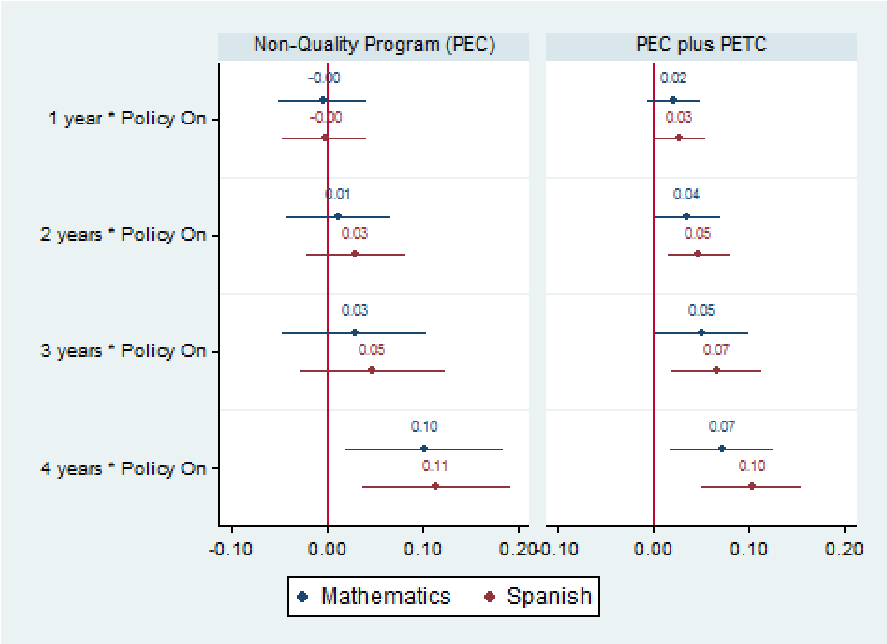


Figure obtained from the point estimators and the 95% confidence intervals coming from a DID regression including a set of dummy variables interacting a post policy dummy with the number of years since intervention. The regression also includes school fixed effects, time-fixed effects and full set of school characteristics as controls. The counter-factual is constructed from all non-PETC schools.

2.5 Impact channels of the effects of PETC on test scores

2.5.1 Does PETC have an effect on students with different abilities?

Table 2.8 shows the effects of a DiD specification on the distribution of math and Spanish scores for PETC schools compared to all non-PETC schools. Columns show the proportion of students graded as insufficient to excellent as reported in ENLACE. The estimations suggest that the overall effect of PETC on math scores comes from a decrease of 2.0 percentage points (pp) in the proportion of students with elementary results combined with an increase of 1.7 pp of those graded as excellent, implying that children at the bottom of the distribution are not benefiting from an increase in the time of instruction. Conversely, PETC results on Spanish seem to have an impact across all the distribution of scores.

TABLE 2.8: Effects of PETC on mathematics and Spanish standardized test scores on the proportion of students graded as insufficient to excellent

	(1) Insufficient	(2) Elementary	(3) Good	(4) Excellent
A. Mathematics				
PETC * Policy On	0.140	-1.991***	-0.204	1.713***
	(0.298)	(0.301)	(0.263)	(0.256)
Number of Schools	164525	164525	164525	164525
R^2	0.166	0.071	0.179	0.107
B. Spanish				
PETC * Policy On	-0.708**	-1.482***	0.777***	1.120***
	(0.294)	(0.297)	(0.278)	(0.177)
Number of Schools	164520	164520	164520	164520
R^2	0.162	0.088	0.131	0.133
Control variables	Yes	Yes	Yes	Yes
School fixed-effects	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Standard errors, clustered on school, in parentheses

Effects on the proportion of pupils graded as insufficient to excellent in mathematics scores conditioned to school's marginality and separated by years since intervention, are presented in Figure 2.5. The point estimators suggest that the higher treatment effects of PETC observed on high marginality schools come, in the beginning, from a significant impact on children at the top of the distribution, but gradually, this effect combines with a reduction in the percentage of children graded as insufficient and elementary. For example, in the case of low marginality schools, the small positive effects revised seem to be driven by children at the top and bottom of the distribution moreover after three and four years of treatment.

Schools with high marginality present a significant increase of 1.2 pp in the proportion of pupils obtaining excellent scores in mathematics one year after policy intervention (i.e an increase of 45% of the base proportion of 3% before policy). More importantly, four years after policy, this proportion exhibits an important growth to 7.2 pp, or 2.7 times the base percentage. This combines with a fall of 3.0 pp in the proportion of students graded as insufficient four years after intervention (i.e. a reduction of 9% to the base proportion of 36%) and 5.8 pp in the percentage of pupils obtaining elementary results (13% of the pre-policy share of 45%).

FIGURE 2.5: Average effects of PETC on the distribution of mathematics standardized test-scores by marginality level and years of treatment

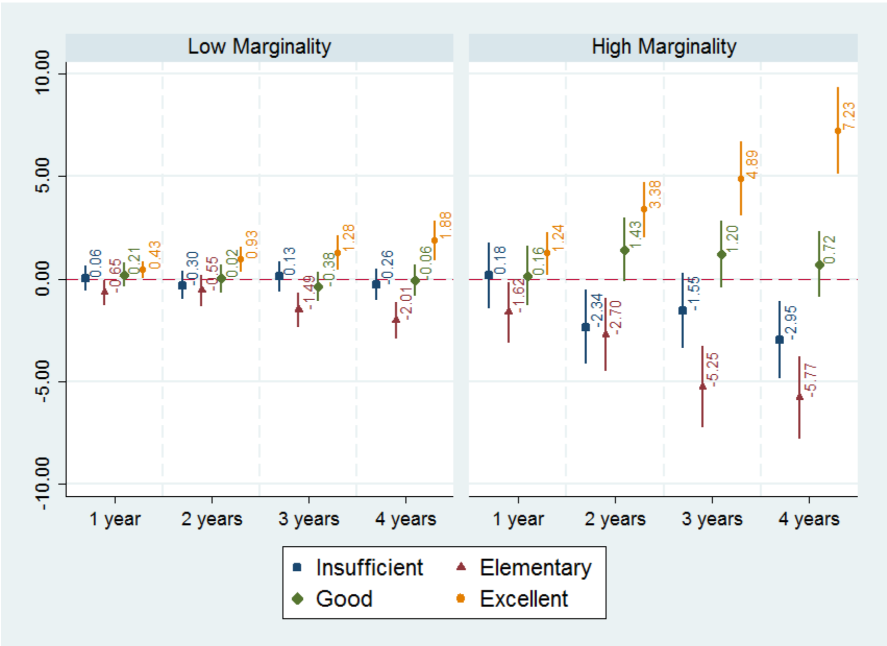


Figure obtained from the point estimators and the 95% confidence intervals coming from a DID regression including a set of dummy variables interacting a post policy dummy with the number of years since intervention. The regression also includes school fixed effects, time-fixed effects and full set of school characteristics as controls. The counter-factual is constructed from all non-PETC schools.

Results for Spanish are presented in Figure 2.6 and suggest a clearer pattern for the most deprived schools, where students with all different type of abilities are impacted from the second year of PETC. For example, the proportion of pupils graded as insufficient and elementary reduces 3.2 pp and 2.1 pp after two years of treatment, respectively, and this reduction grows to 4.8 pp and 4.0 pp four years after policy, representing a decrease of 14% respect to the base proportion of 36% in the case of children graded as insufficient and a smaller 8% respect to the 48% of pupils graded as elementary before policy intervention.

At the top of the distribution there is a significant increase of 3.0 pp in the proportion of students obtaining good grades and 2.0 pp for those with excellent results and these effects grow after four years of treatment to 4.2 pp and 3.7 pp, respectively. This represents, four years after intervention, a change of 30% in the proportion of students with good results respect to the base proportion of 14% before PETC. Similarly for the case of students graded as excellent there is an increase of 2.5 times the base proportion of 1.5%.

FIGURE 2.6: Average effects of PETC on the distribution of Spanish standardized test-scores by marginality level and years of treatment

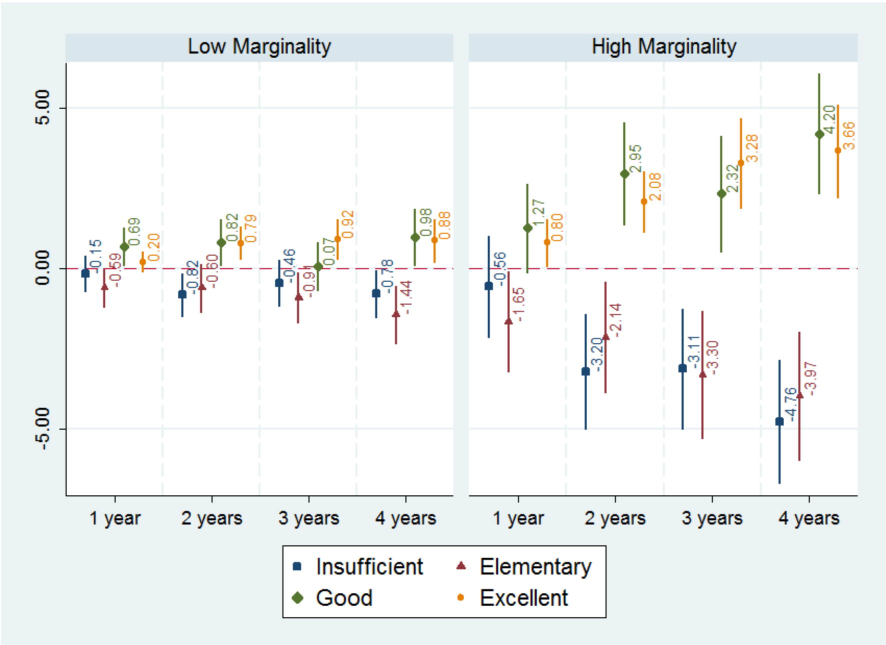


Figure obtained from the point estimators and the 95% confidence intervals coming from a DID regression including a set of dummy variables interacting a post policy dummy with the number of years since intervention. The regression also includes school fixed effects, time-fixed effects and full set of school characteristics as controls. The counter-factual is constructed from all non-PETC schools.

Jointly these results suggest that language skills are absorbed in the mid-run by students with different abilities within PETC schools, moreover with a lower socioeconomic environment.This evidence can be interpreted as mechanism that could indeed reduce differences between disadvantaged and more advantaged pupils within high marginality schools. Nonetheless, for the case of math, since the higher effects in more deprived schools are apparently explained by an important push of children at the top and bottom of the distribution of scores, it is not clear that the program is reducing differences between the “best’ and “worst” math students in PETC schools across time. Nonetheless, according to the overall results conditioned on school’s marginality, it is clear that a reduction in the gap between deprived and advantaged schools is taking place.

However, a major concern arises from the reduction in the proportion of pupils graded as insufficient and elementary in both subjects, since this may well be explained by students simply stepping out of schools. It is plausible to think that longer school days are harder to cope by those with lower abilities and in more deprived areas. Drop out rates in Mexico are nowadays rather low in primary education (1.9% in the period here analyzed according to the Statistics 911) but in order to address any concern regarding the effects of PETC on desertion, Table 2.9 shows the effect of the intervention on dropout rates in schools which present desertion at any given grade and year, on average and by level of marginality.[21] The results suggest that desertion is not driven or modified by the presence of the policy neither on average nor in more or less deprived schools.[22]

TABLE 2.9: Effects of PETC on dropout rates by level of marginality

	(1)	(2)	(3)	(4)	(5)	(6)
Control Schools	All non-PETC schools			With Matching		
	Average	Low Marg.	High Marg.	Average	Low Marg.	High Marg.
PETC * Policy On	0.187	-0.050	0.249	0.046	-0.027	0.032
	(0.123)	(0.143)	(0.245)	(0.131)	(0.183)	(0.280)
Number of Schools	154989	80356	74633	51921	34543	17378
R^2	0.097	0.120	0.075	0.098	0.105	0.080
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
School fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Coumns 1 to 3 show standard errors, clustered on school, in parentheses. Columns 4 to 6 show bootstrap standard errors with 100% of replacement and 100 repetitions, clustered on school, in parentheses.

2.5.2 Are PETC effects driven by a selection of students?

As discussed, one of the main points raised by teachers and school principals in the qualitative evaluation conducted by UNESCO (2010), is the increase parent's demand for full-time schools. A worrying concern surging from a higher demand of PETC schools is that principals and teachers may have more room to select best new students, who would on average present better results in standardized tests.[23] Consequently, the positive results of the program as discussed before, may well be explained by selection rather than policy intervention.

The results presented so far do not support this hypothesis, given that schools selection (or student's self-selection into PETC schools) can only happen for newcomers who cannot replace other students already registered at school, and the program is having an impact not only on students with higher scores but also on children at the bottom of

[21] Schools that present a positive inflow of students are analyzed separately below.
[22] Placebo tests on dropout rates are presented in Table B6 in the Appendix
[23] Of course there is also the possibility of auto-selection where new students can be more motivated than the average, since conceivably, most motivated parents would be those looking to move their children from a non-PETC to a PETC school.

the scores distribution. Selection may explain gains in the upper part of the distribution of test scores, but it is more difficult to think of a mechanism for which it could have an effect on those more behind who are also showing improvements. Furthermore, had the positive impact been explained by pure selection, one would expect low marginality schools to have a higher chance to select "better" students, and possibly have stronger average impacts than high marginality schools, and this is not the case supported by the evidence.

Finally, given that primary schools in Mexico cannot dismiss students already registered, if there is a mechanism acting to select "better" or more motivated students in order to achieve higher results in ENLACE, the proportion of newcomers in PETC schools should have an effect on test scores. In this regard, Table 2.10 shows the results of a school and time fixed-effects model on test scores including a set of controls and separated by level of marginality. Estimations are in general significant but very close to zero indicating that the proportion of new students at any given grade and year in PETC schools are not positively influencing test scores. Hence, PETC effects are plausibly not driven by selection.

TABLE 2.10: Fixed-effects OLS of the proportion of new students at any given grade and year in PETC Schools on Spanish and mathematics test scores

	(1)	(2)	(3)
Subject	Average	Low Marg	High Marg
A. Mathematics			
Proportion of new students	-0.005**	-0.005**	-0.002
	(0.002)	(0.002)	(0.008)
Number of Schools	5512	4388	1124
R^2	0.716	0.737	0.725
B. Spanish			
Proportion of new students	-0.005**	-0.005**	-0.004
	(0.002)	(0.002)	(0.007)
Number of Schools	5511	4388	1123
R^2	0.716	0.740	0.710
Control variables	Yes	Yes	Yes
School fixed-effects	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes

Standard errors, clustered on school, in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.6 Conclusions

This work analyzes the potential effect on pupil performance in Mexican primary schools of a change in the time of instruction from 4.5 to 8 hours, the inclusion of new pedagogic tools used for children enrolled in these schools, and the structure of teaching implemented by the Full-Time Schools Program (PETC). The gradual inclusion of schools in the program allowed for the construction of four treatment and control groups as a natural experiment investigating what happened to pupil achievement in schools were the policy was introduced relative to pupils in schools that were not subject to PETC during the whole period. Additionally, this is compared to a matched control group. Hence, DiD and PSM plus DiD regressions were conducted to conclude overall effects of the policy separated by years of treatment and school's marginality and to study effects on kids with different abilities.

After providing evidence that suggests there are no differences in the trends of pupil test scores in PETC schools relative to comparison schools in the pre-policy period, effects on Spanish and mathematics scores exhibit a significant and positive effect on both subjects. The precise impact ranges from 0.05 SD after two years of treatment to 0.11 SD after 4 years of treatment on both subjects using a panel of schools with a full set of controls. These effects are arguably robust to the application of 'placebo tests', examination of different treatment and control groups and the matching of control schools with similar observable characteristics.

The results also show a stronger impact on average in schools with high marginality compared to less deprived schools. DiD results show an effect of at least 0.12 SD after two years of treatment and of 0.29 SD after four years of treatment on both subjects. These results compare to non-significant average effects on low marginality schools during the first three years of treatment and a lower positive effect four years after intervention of around 0.05 SD and 0.07 SD in math and Spanish, respectively. The fact that high marginality schools are getting the best results signifies a reduction in the gap between less and more advantaged schools to a half in math and in a third for the case of Spanish test scores.

After inspecting PETC effects on the distribution of scores results suggest that in the case of mathematics, after four years of treatment there is a clear pattern of a reduction in the proportion of students graded as insufficient and an increase of those with excellent results. This pattern is observed more clearly for schools with high marginality. For the case of Spanish, policy intervention exhibits effects across all the distribution of scores also with stronger impacts on high marginality schools. These results are of key relevance to highlight that less skilled kids even in deprived environments, are also benefiting for longer school days.

Further inspections conducted on causal channels show that the program does not have an effect on drop out rates emphasizing the fact that low achievement students are indeed benefiting from this policy. Finally, the proportion of new students in treated schools does not have a positive effect on test scores, allowing to argue against selection of "better" students as the mechanism for which PETC is having showing improvements at the top of the distribution of test scores.

Despite PETC schools treated 2013 present significant differences in Spanish test scores between treated and controls three years before treatment (as show in the placebo tests), there are no significant differences for all treated groups one and two years before policy introduction, giving a good support for causal inference. Having subjected the identification strategy to a number of robustness checks including the generation of a smaller control group with similar observable characteristics to the treated, results should constitute a PETC effect on test scores.

The overall findings of this research are of considerable significance when placed into the wider education debate about what works best in schools for improving pupil performance. Despite the average gain in test scores for PETC schools is relatively small on average, they are in line with the findings for other Latin-American programs of a change in the instruction time in basic schools. More importantly, findings on the impact of PETC schools are sustained four years after policy intervention and are higher in more deprived schools compared to those found in comparable programs in the region.

Chapter 3

Leave them kids alone! The positive effect of abolishing grade retention on pupils' dropout rates: Evidence from a policy change

Automatic promotion of students who fail to meet specified promotion criteria is often opposed to on the grounds that it lowers school expectations and student achievement. In some school systems, grade repetition (sometimes referred as grade retention) is seen as a valid corrective measure for underachievers as it may work as a deterrent to poor school performance by inflicting a high penalty on under-performers, working as an incentive for students to increase their efforts (Manacorda, 2012). Supporters of this policy also argue that retaining children helps their maturation and it gives them time to meet the minimum academic standards of any given grade. Indeed, some evidence in developed countries has attributed positive effects of retention on academic achievement (Jacob and Lefgren, 2004; Dong, 2010).

On the other hand, grade repetition may be harmful for pupils, as it raises costs for schools and families, affects children's motivation and self-esteem, stigmatizes children and increases the likelihood of dropout. Some authors argue that those who are retained may be at a higher risk of dropping out because pupils who are older than their classmates may feel different than their peers and discouraged (Holmes et al., 1989; Roderick, 1994). Furthermore, it is considered that repeating grades delays entrance of students into the labor market which may pose substantial monetary cost on individuals over the life-cycle. Consequently, some schools are experimenting with different alternatives to grade repetition in favor of other types of ”social promotion” for children who are not keeping up, involving a more holistic evaluation considering parent’s participation and remedial academic assistance.[1] [2]

Studies in different areas, including economics, psychology and sociology, have broadly documented that pupils who repeat a grade do not improve their school outcomes and are significantly more likely to dropout compared to continuously promoted students (Roderick, 1994; Gomes-Neto and Hanushek, 1994; Jimerson, Anderson, and Whipple, 2002; Manacorda, 2012). Nonetheless, there is a less well studied causal relationship between retention in grade and its possible outcomes, including dropout, regardless of children’s Socioeconomic Status (SES) and in contexts of high educational privation as observed in low- and mid-income countries.

Considering the lack of conclusive empirical findings on the effects of repeaters in poorer countries, the use of public resources to reduce repetition rates is not completely justified. In effect, this paper seeks to evaluate the effects of abolishing grade retention in the first three grades of primary school on pupil’s dropout rates in the context of a developing country with universal primary education coverage. In August 2012, Mexico passed from a punitive system in which all children graded below a threshold were not promoted to the next academic level, to one focused on the mandatory ”social promotion” of students. The reform allowed children in the first three years of primary education to pass to the next grade regardless of their level of achievement. This change caused an exogenous reduction independent of school’s SES of grade repetition rates in primary schools with varying trends before policy reform. This sharp effect in a short-span of time, allows to study the effects of such policy in a novel quasi-experimental manner using a panel of schools and administrative data from 2007 to 2014.

Definitely, neither automatic promotion nor grade repetition completely address the problems of low achievers, as there are different reasons to repeat a grade and some are even voluntary. For example, a family may consider this necessary for their offspring

[1] For a revision of common alternatives to retention in-grade, see Protheroe (2007).

[2] This also relates to cross-national variations associated with contrasting systems of schooling where historical reasons also take play. For example, Brophy (2006) documents that Scandinavia and the English-speaking countries (and developing countries influenced by them) emphasize universal education to higher levels, encourage social promotion and grade repetition rates are low. Whereas, France, Portugal, and Spain (and developing countries influenced by them) emphasize universal education at lower levels but limit admittance to secondary and post-secondary levels, so repetition rates are higher in lower school levels.

to fully acquire the knowledge from an specific grade. Retention in grade is "rational" if for example, children are continuously absent from school to work or help their family in housekeeping tasks and this is clearly plausible in poorer countries or contexts (Gomes-Neto and Hanushek, 1994). Similarly, other factors outside schools may as well determine student's involuntary grade retention and most of them also relate to their socioeconomic status (SES). High direct costs for example, for buying uniforms, writing materials, textbooks, etc., and sensitivity to the opportunity costs of attending school are more likely to strike the children from impoverished backgrounds. Undoubtedly, parent's education and cultural capital have an effect on both, children's repetition and dropout likelihood. Finally, malnutrition, which is also related to children's SES, may also affect their achievement (Pollitt et al., 1998). In any case, grade repetition is costly for families and schools, since it increases school needs of material and human resources and family pecuniary and non-pecuniary investments.

Disentangling socioeconomic and school factors from grade repetition and dropout rates is difficult in a non-experimental set-up. Poor families and teachers and low-schools quality, teachers absenteeism, and lack of school infrastructures, often cited as major problems of school systems in developing countries, may explain both high repetition rates and students incentives to abandon school. Therefore, it remains important to understand the practical consequences of systems and policies seeking to abolish retention in grade in favor of a more cost-efficient "social promotion", as they may be undone if for example, children's socioeconomic context and family decisions are the main source of high repetition rates. In this case, other policies directed to eliminate the economic and context causes of grade repetition rather than repetition itself may be more effective.[3] Similarly, untended effects of the policy may have disincentives on children's performance at school. As, for example, evidence on Brazil has shown a negative and significant causal effect of automatic promotion on math test scores Koppensteiner (2014).

The estimations presented in this chapter show a positive effect of exogenously reducing grade repetition on dropout rates despite the short period of analysis. Two-way fixed effects models exhibit an impact of 0.3 % points lower dropout rates after policy change. Results are consistent across different socioeconomic contexts and remain after two years in either, public and private institutions and in schools with different "family investments". Furthermore, regressions on the impacts of policy change on student's

[3] For example, Patrinos, López-Calva, Bando et al. (2005) show evidence that in Mexican indigenous households, conditional cash transfers coming from the PROGRESA program have increased attendance and reduced grade repetition and school dropout rates.

achievement show that eliminating the ”threat” of grade retention does not have negative effects on the average results of standardized Spanish and mathematics test scores. These results are a clear contribution on the broader debate in favor of the ”social promotion” of children in less developed countries.

The rest of this chapter is structured as follows: Section 3.1 discusses in detail policy change in promotion criteria in Mexico. Section 3.2 presents the panel of schools used for the analysis and some descriptive statistics. Section 3.3 describes the empirical strategy. Section 3.4 discuses average effects and presents further robustness checks. Section 3.4 shows the heterogeneous effects of the reform in different types of schools. Section 3.6 includes some concluding remarks.

3.1 Automatic promotion in Mexican Schools

On August 2012 the Mexican Secretariat of Education (SEP, for its abbreviation in Spanish) conducted a reform of the general rules for the student's evaluation, promotion and certification in basic education starting in the academic cycle 2012-2013.[4] Before the reform, all children in primary and secondary schools with an average mark lower than six (on a scale of one to ten), were considered to have an 'insufficient' level of knowledge to be promoted to the next grade or educational level. Nonetheless, the change in grade retention laws transformed the way children from first to third grade of primary education were evaluated and promoted. As from the academic cycle 2012-2013, all first to third graders in both, public and private schools would be automatically promoted to the next grade. The reform at the national level, also established that first to third graders who got a mark lower than six would only repeat grade if their parent's required it. [5]

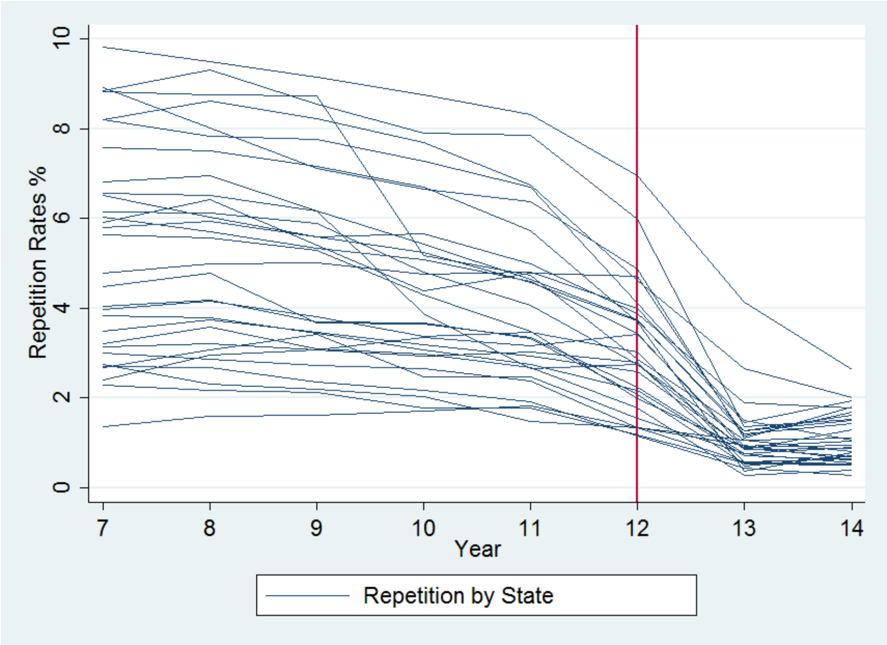
Figure 3.1 shows grade repetition trends in first to third grades for the 31 Mexican States and the Federal District. The lines show a sharp reduction in grade repetition from varying higher rates to varying lower rates, generally converging below a 1% of repetition and remaining stable through the next academic cycle,[6] Except for the case of two states with traditional lower educational outcomes, Oaxaca and Michoacán for which repetition rates remained notably higher than the average. Note that the nature of the intervention caused a higher/ lower impact in States (and schools) with different repetition levels in the pre-policy period, allowing to compare some schools with higher repetition to others with lower rates before and after the exogenous change in policy. This is the key argument for our identification strategy as it will be discussed in Subsection 3.3.

[4] All details of the reform can be consulted in the 'Pact 648' (Acuerdo 648 in Spanish) in the Nation's Official Diary (Diario Oficial de la Nación) published on August 14th, 2012.

[5] It is worth mentioning that before policy change, all children were graded using evaluations designed by each teacher. Thus it cannot be ruled out that teacher's criteria had an influence on children's retention in-grade. However, despite the lack of a national test to evaluate whether children are promoted to the next grade or not, there exist a national curricula and a national criteria of what should be the standard knowledge acquired at any grade during basic education in Mexico, including primary and secondary schools. Hence teacher's design of evaluations and criteria should not be the dominant cause of the variation in school's average grade repetition rates before policy change. For the purpose of this research, all the results presented in subsequent sections include controls for teacher's education as a proxy of their characteristics in the classroom.

[6] Note that despite the policy change abolished retention in-grade, the remaining 1% of repetition rates may be explained by pearents requiring their children to repeat a grade, for different own personal reasons involving maturation, financial restrictions, etcetera.

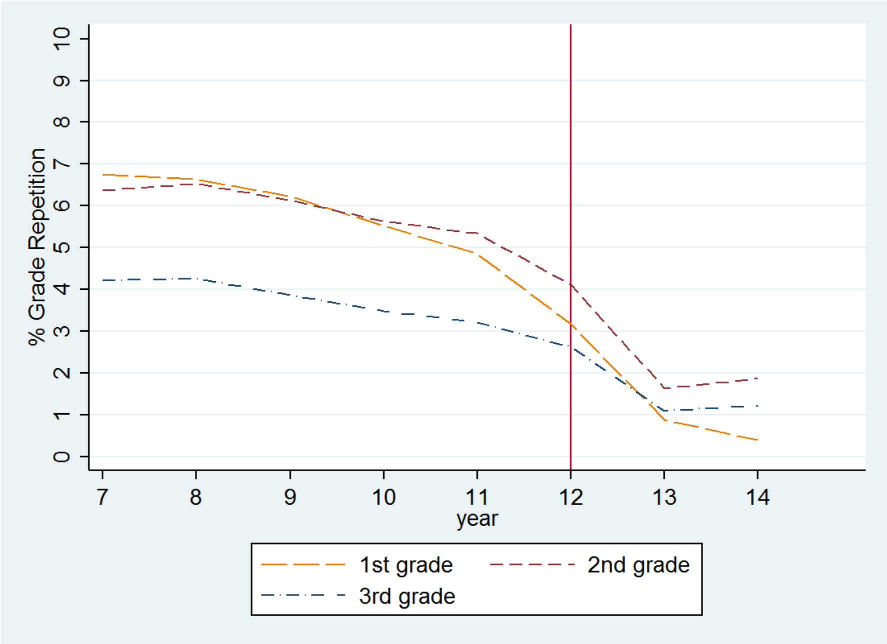
FIGURE 3.1: Average repetition rates in 1st to 3rd grades: trends by States before and after policy change



In September 2013, the newly elected Mexican government made an amendment to the recently approved general rules for the evaluation of children. The new reform, approved only some days before the next academic cycle started, brought back the 'conditional' possibility of grade repetition for second and third graders. Hence, only children in the first year of primary education would be automatically promoted to the next year regardless of their marks. However with the newest amendment, even those children in second and third grade who had 'insufficient' results, could be 'conditionally' promoted according to teacher's personal criteria and 'parent's commitment' to support children's tasks in the next academic year. Additionally, children could only repeat a year once, either in the second or third grade. This way, if a children was retained in second grade, by no means she could be retained in third grade. Moreover, teachers should now grade children not only based on exams but the also on "specif children's needs and their context".[7] In practical terms, the abolition of the reform didn't reverse the changes previously applied and did not affect considerably repetition rates in second and third grade in the next academic year 2013-2014 as shown in Figure 3.2. For which, we assume that the effect of abolishing retention in grade, from first to third grade of primary education, remained active since September 2012 and throughout the next two academic cycles.

[7] All details about the 2013 reform can be consulted in the 'Pact 696' (Acuerdo 696 in Spanish) in the Nation's Official Diary (Diario Oficial de la Nación) published on September 30th, 2013.

FIGURE 3.2: Average repetition rates in 1st, 2nd and 3rd grades before and after policy



3.2 Data and Descriptive Statistics

The data for this paper comes primarily from the annual national *Statistics 911* in the academic years 2006-2007 to 2013-2014. This is a census dataset coming from schools questionnaires applied by SEP at the beginning and at the end of each academic year. The *Statistics 911* offer information on school's infrastructure such as number of classrooms and basic services like water and electricity. They also offer characteristics of teachers and principals like the proportion of those in each school with any given level of education, from primary to the postgraduate level, and information on number of students by grade, age and sex, number of pupils who passed and failed and those who are repeating a grade. This allows to estimate the total proportion of children repeating grade in any given year and those dropping out from school during the academic year (intra) and between school cycles (inter) and consequently the total dropout rate, considered as the sum of intra- and inter-course dropout rates.

Note that the *Statistics 911* only offer information at the school level, for which our dropout measure represents changes in net enrollment at school j, in time t for grade g. Consequently, we are not able to observe when any specific child drops out from school, but we can realize when a school "loses" students in any given grade at the end of each academic year and at the beginning of the next one. Naturally, some of these children dropping out from a school may move to a different one, however, on average we can still compute an overall measure of dropouts. For example, if every school is "losing" less/more children in any given grade g in time t with respect to $t-1$, we assume that the average dropout rates for grade g in time t are lower/higher. In other words, we can observe if the school system as a whole is "losing" more or less students across time. Effectively, as we cannot observe which children are simply changing schools and which are dropping out from the school system, the scope of this research is limited to the analysis of average changes to the net enrollment in primary schools.[8]

Statistics 911 are also merged with each schools' proportion of students receiving a PROSPERA Scholarship[9] in order to explore differences between poorer and richer schools. Furthermore, we also integrate the average third grade results in standardized tests of Spanish and mathematics from 2008 to 2013 to test if abolishing the 'threat' of grade repetition had an impact in student's achievement in that grade.[10] ENLACE includes the number of students tested and the proportion of unreliable tests in each school.[11] It also reports five categories of 'privation' suffered in school's localities, which work to build a variable indicating if schools are from a 'low' socioeconomic status (SES) or a medium- to high-SES otherwise.[12]

The main descriptive statistics are presented in Table 3.1. We exclude indigenous schools from the sample (about 7% of the total) because these are located in remote

[8]In specific, our dropout measures indicate the inter-dropout rates consisting of those students abandoning their courses between two academic years t and $t+1$ for every 100 students enrolled at the beginning of the academic year t, as indicated by the formula: $(({}_{t}FE^{g} - {}_{t+1}IE^{g+1}) - ({}_{t}\ SF^{g+1})/{}_{t}FE^{g}) * 100$, where ${}_{t}FE^{g}$ represents students enrollment at the end of the academic cycle t in grade g; ${}_{t+1}IE^{g+1}$ represents students enrollment in the next grade $g+1$ at the beginning of the next academic cycle $t+1$; and ${}_{t}\ SF^{g+1}$ represents the number of students who failed grade $g+1$ in time t. We also compute intra-dropout rates according to the formula $(tFE^{g} - tIE^{g})/100$, that considers changes between initial and final enrollment during the academic cycle t in grade g. Consequently, the total dropout rate is given by the sum of both the inter- and intra-dropout measures.

[9]PROSPERA, before PROGRESA or OPORTUNIDADES, is a conditional cash transfer program directed to the poorest families in Mexico.

[10]ENLACE test was suspended by the federal authorities after the academic cycle 2012-2013. A new test named PLANEA was applied for the school cycle 2014-2015 but statistics on dropout rates and grade repetition are still not available for such period.

[11]Every year a set of questions to be used in the next year's test is applied to a controlled sample, this works to built the standardized scale of the next year's test and allows to identify students out of this scale who are labeled as unreliable. Furthermore, ENLACE includes quality controls through an automatic validation to detect collusion with the use of the models K-Index and Scrutiny as described in technical details of the ENLACE manual.

[12]The five different categories of privation are provided by the National Council of Population (CONAPO, for its abbreviation in Spanish). These are based in eight socioeconomic variables of the school's locality: average education levels, household's characteristics (i.e. available services and infrastructure) and goods availability. Low-SES are those schools with the two highest levels of privation whereas mid- to high-SES are those located in the lowest three categories. For further details see CONAPO (2010).

areas, they are often multi-grade and thus, grade repetition is harder to compute. Therefore we keep public and private general schools. The number of units for which there exists full information about grade repetition and dropout rates for the eight years period in a balanced panel is of 463,424 school-year observations or approximately 59,000 schools in each academic year. This represents approximately 77% of the total private and public schools that exist in Mexico. Note that, regarding ENLACE test scores, the number of observations available is lower, this is because we only include information from 2008-2009 to 2012-2013. Average grade retention rates for the cited period corresponding to 1st to 3rd grades is of 3.82% while total dropout rates are of 1.51% on average. Dropout rates are higher when we consider all primary education (1.93%) while retention is lower (2.62%). This implies that grade repetition is higher in lower grades, when children are younger, but dropout is higher in 4th to 6th grades when students are typically between 9 and 12 years old.

TABLE 3.1: Main Descriptive Statistics

Variable	N*Years	Mean	SD	Min	Max
Grade Repetition 1st-3rd	$57,928*8$	3·82	5·22	0	100
Dropout 1st-3rd	$57,928*8$	1·51	10·88	−100	100
Repetition 1st-6th	$57,928*8$	2·62	3·59	0	100
Dropout 1st-6th	$57,919*8$	1·93	9·01	−100	100
A. Teacher's Education					
Vocational	$57,850*8$	37·05	33·00	0	100
Bachelors	$57,850*8$	56·79	33·44	0	100
Postgraduate	$57,850*8$	5·20	12·58	0	100
B. Schools's Characteristics					
Number of Students	$57,928*8$	213·69	174·27	20	1416
Number of Teachers	$57,928*8$	7·64	5·07	1	30
"Carrera Magisterial"	$57,928*8$	3·15	3·75	0	33
Low-SES School	$57,928*8$	0·35	0·48	0	1
Public School	$57,928*8$	0·89	0·31	0	1
Morning Shift	$57,928*8$	0·82	0·38	0	1
C. ENLACE results					
Math Avg. 3rd Grade	$62,103*6$	535·38	75·82	103	914
Spanish Avg. 3rd Grade	$62,101*6$	538·44	82·25	64	897

Source: Author's elaboration using Statistics 911 for the school years 2006-2007 to 2013-2014 and ENLACE databases from 2008-2009 to 2012-2013. "Carrera Magisterial" is a national program that offers monetary incentives to "the best" teachers according to a set of standardized evaluations. The variable presented refers to the percentage of teachers receiving such incentives in each school.

TABLE 3.2: Average repetition and dropout rates from 2007 to 2014

	Grade Repetition		Dropout	
School Year	*1st-3rd Grades*	*1st-6th Grades*	*1st-3rd Grades*	*1st-6th Grades*
2006-2007	5.79	3.95	3.14	3.25
2007-2008	5.83	4.01	2.59	2.71
2008-2009	5.41	3.75	2.53	2.56
2009-2010	4.88	3.32	1.82	1.94
2010-2011	4.46	3.01	1.76	1.72
2011-2012	3.26	2.22	1.03	1.29
2012-2013	1.18	0.92	0.87	0.94
2013-2014	1.14	0.91	0.06	0.61

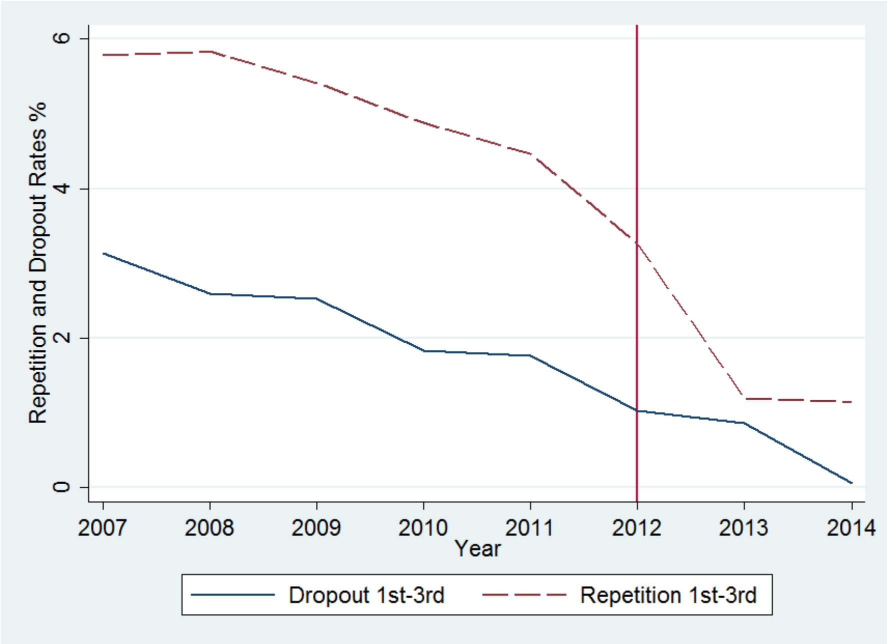
Source: Author's elaboration using Statistics 911 for the school years 2006-2007 to 2013-2014

Table 3.2 presents average grade repetition rates for 1st to 3rd and 1th to 6th grades across years. Note the decreasing trend for all variables along with a clear sharp reduction in both grade repetition and dropout rates from the school cycle 2012-2013. For example in the case of 1st to 3rd grades, we observe a decline in retention rates, from 3.26% to 1.14% from 2011-2012 to 2013-2014, respectively. This same reduction is registered in dropout rates, as they go from an already low 1.03% to a lower 0.06%, respectively. Such trends can be graphically depicted in Figure 3.3. It is worth mentioning again the clear downwards pre-trend in both, repetition and dropout rates. A sense of anticipation to the policy change may be suggested, however, as mentioned, repetition laws in Mexico are only approved at the national level and it would be unlawful if some states or schools change their evaluation criteria without a previous consultation to the federal authorities. In any case, any state/county pretrend is addressed in the methodology and falsification tests presented in the coming sections.

3.3 Empirical Specification

Our estimation strategy uses the variation in grade repetition rates across schools in time to identify the effect on dropout rates of abolishing grade retention for students not passing mandated assessments. The intuition is that schools with higher grade repetition rates would be affected 'more intensively' by the exogenous change in policy, compared to schools with already low levels of grade repetition. The key assumption is that the reduction in grade repetition rates is independent of unobserved school characteristics. This is plausible, as the elimination of retention in grade was mandatory for all

FIGURE 3.3: Average grade repetition and dropout rate in 1st to 3rd grades



schools regardless of their characteristics. As a matter of fact, Figures 3.4 and 3.5 show that regardless of school's socioeconomic status (SES) and type of funding (public or private) all schools, after policy change, converge towards the same repetition rate of around 1%.[13] The clear downwards trend in retention rates presented for all states may reflect the general improvement in education outcomes across different States in Mexico through time. However, there is the concern that it may also represent an 'anticipation' from local authorities to the policy change. To the best of our knowledge, States in Mexico cannot change promotion rules as these are determined by law at the federal level for which "social promotion" was not allowed before 2012. Additionally, different time trends at the state and county levels are considered in the corresponding estimations to account for differences between state or county in the general downward trend across schools explained by unobserved factors happening at the county/state level. Thus if lower retention is associated with lower dropout rates the change in grade repetition rates from 2012 to 2013 will be negatively related to the proportion of children abandoning school after the reform. This is captured by the following equation:

$$Dropout_{sti} = \alpha_1 + \alpha_2 Repetition_{sti} * Post + D_t + D_s + \alpha_3 X_{st} + D_j * Year + \varepsilon_{stij} \quad (3.1)$$

[13]Despite policy change completely abolished grade retention, the rates do not completely go to zero. This may be due to children abandoning school during the school year and coming back one year later to repeat the grade in which they were registered. Another reason may be voluntary repetition, when parent's believe that their children did not accumulate the 'necessary' knowledge for any given grade.

Where $Dropout_{sti}$ is the total dropout rate in school s in time t=2007-2014 and grade i.'Post' is a dummy taking the value of 1, in the after policy period and zero otherwise. $Repetition_{sti}$ measures average repetition rates in the pre-policy period and after policy change in school s and grade i. Hence, $Repetition_{sti} * Post$ captures the effect of the exogenous change in repetition rates for school s in time t=2013/2014. X_{st} includes school level potential predictors of dropout rates for the whole pre- and post-policy period, like number of students, teachers per every 100 pupils and teachers' and principals' average education. D_t is a time dummy variable that accounts for changes in pre- and post-policy period that might have affected school's outcome trends differently. D_s are time invariant school dummies capturing fixed school characteristics which might affect the outcome of interest. Finally, D_j*Year represents state j (or county j) year trends that may correlate to policy change.

This two-way fixed-effects model compares schools dropout rates between units with high and low repetition rates before and after policy change. The unit of observation is schools but as we include county/state year trends we cluster standard errors on state level. α_2 is our coefficient of interest and measures the impact of abolishing retention in-grade (or exogenously reducing grade repetition rates) on drop out rates in percentage points.[14] If the policy change had an impact we would expect α_2 to be negative. More specifically, this term will tell us for schools where repetition rates are 1% point higher in the pre-policy period, how much would dropout rates change after abolishing grade retention.

The causal interpretation of α_2 in the above framework rests on the assumption that after controlling for time fixed-effects, state/county year trends and time invariant characteristics of schools, ε_{tsij} is independent of the interaction term. In other words, that there are no time-varying unobserved school-specific and state/county factors that are correlated with the change in grade repetition after policy intervention. As mentioned, we believe this is rather plausible as the sharp reduction in retention in-grade comes from a modification of the federal law to asses pupils independent of school characteristics.

In section 3.5 we also present different estimations of α_2 on inter-course and intra-course dropout rates. We believe that the nature of this two measures is different. For example, inter-course dropout rates take into account all those children who finish grade i in year t but do not come back to school the next year. Such children (and their parents) could be discouraged by the grades they receive at the end of the year increasing

[14]This strategy is similar to the one used by (Duflo, 2000) and more recently by Chakraborty and Bakshi (2016)

their likelihood to abandon school. On the other hand, children abandoning school before even knowing their grades (intra-course) could be more driven by other factors independent of their performance and more related to their family's socioeconomic and cultural conditions which are not affected in any way by a policy for the "social promotion" of children. Consequently, we expect that the change in retention rules should affect more inter-course dropout rates, as children are not failed anymore regardless of their grades, and thus, they are plausibly "encouraged" to continue in school.

In this regard, we also present different estimations of α_2 for subgroups of schools according to their type (public or private), their SES and a proxy of varying "family investments" on children. The idea is to test if the social promotion of pupils is encouraging them to continue in school after "failing" a grade, regardless of their context. In such case, we expect inter-course dropout rates to show a reduction after policy change regardless of pupil's socioeconomic status, while intra-dropout rates, which are on average higher in poorer contexts, should not be strongly affected. With this, we seek to deepen the analysis on the nature of dropout rates and if they depend entirely on children's socioeconomic and family context or if there is indeed a chance for school policies and in specific for "social promotion" to reduce them.

FIGURE 3.4: Average grade repetition 1st to 3rd grade: trends before and after policy change by SES

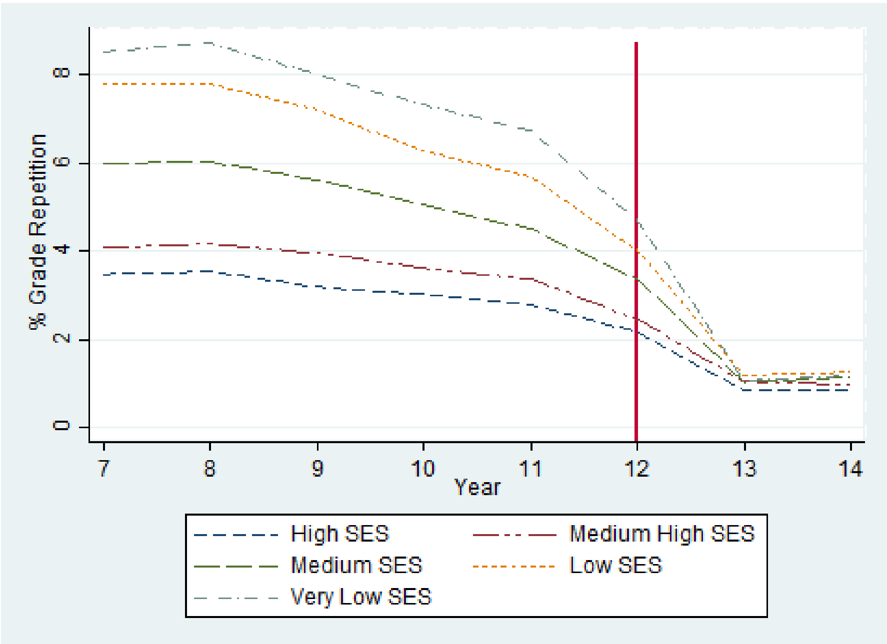
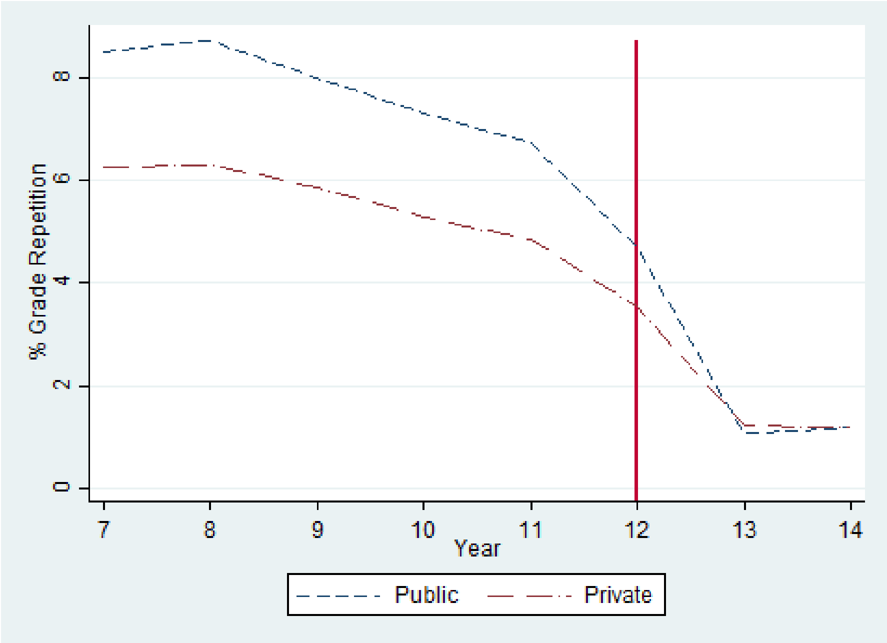


FIGURE 3.5: Average grade repetition in 1st to 3rd grades of public and private schools: Trends before and after policy change



3.4 Does abolishing retention in-grade reduce pupil's dropout rates?

Table 3.3 shows the main results of the specification in Equation (1). Column 1 shows the results of the interaction between grade repetition rates and the after policy period, or the effect of the policy change on total dropout rates including school and time fixed effects. Columns 2 and 3 show the effect on total dropout rates by adding state time year fixed-effects and county times year fixed-effects, respectively. For the three different specifications the coefficient attached to the change in policy shows a highly significant reduction in total dropout rates of approximately 0.26% points on average (S.E. of 0.02) for schools with repetition rates 1% higher in the pre-policy period. For which, an school going from 4% to 1% after policy change would translate into approximately a reduction of 0.78 % points in dropout rates. Similarly, considering that the average repetition rate before policy was already at a low 1.18% this impact translates into an average reduction of approximately 21% of the pre-policy change dropout rate. Note that other variables included in the model also seem to present sensible signs regarding their effects on dropouts. For example, bigger schools have more dropouts but increasing the ratio of teachers in 1 per 100 students significantly reduces dropout rates. Not surprisingly, teacher's education at higher levels than vocational studies, also relates negatively to dropout rates.

TABLE 3.3: Two-way fixed-effects models of dropout rates on grade repetition

	Dropout (1)	Dropout (2)	Dropout (3)
Repetition * After	−0·251∗∗∗	−0·269∗∗∗	−0·254∗∗∗
	(0·040)	(0·044)	(0·045)
# Students	0·182∗∗∗	0·183∗∗∗	0·192∗∗∗
	(0·009)	(0·008)	(0·009)
Students-sq	−0·000∗∗∗	−0·000∗∗∗	−0·000∗∗∗
	(0·000)	(0·000)	(0·000)
Teacher/Stud. Ratio	−1·529∗∗∗	−1·545∗∗∗	−1·584∗∗∗
	(0·100)	(0·094)	(0·097)
A. Teacher's Education			
Vocational	−0·007	−0·007	−0·007
	(0·004)	(0·004)	(0·005)
Bachelors	−0·011∗∗	−0·011∗∗	−0·010∗∗
	(0·004)	(0·005)	(0·005)
Postgrad	−0·016∗∗∗	−0·017∗∗∗	−0·016∗∗∗
	(0·005)	(0·005)	(0·005)
Carrera Magisterial	−0·047∗∗	−0·036∗	−0·044∗∗
	(0·020)	(0·020)	(0·020)
School Fixed- Effects	Yes	Yes	Yes
Time Fixed-Effects	Yes	Yes	Yes
State * Year Fixed-Effects	Yes	Yes	No
County * Year Fixed-Effects	Yes	No	Yes
R-squared	0·243	0·259	0·267
Number of schools	435822	435822	435822

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, clustered on state, in parenthesis. Teacher's education is measured as the proportion of these with any given educational level in each school. "Carrera Magisterial" is a national program that offers monetary incentives to "the best" teachers according to a set of standardized evaluations; the variable included in the model refers to the proportion of teachers receiving such incentives in each school. Teachers/Stud Ratio is measured as the number of teachers per 100 students at the school level.

3.4.1 Falsification Tests

Figure 3.6 plots the results of different regressions including interactions between grade repetition and a set of dummies taking the value of 1 in the years 2008-2012, as if policy had changed before 2012. Across all models, we use 2007 repetition rates as a base. Estimations, as those coming from Equation 1 include time, school fixed-effects (left hand figure) and a combination of school, time and county-year fixed-effects (right hand figure). Regressions also include a set of school and teacher's characteristics as controls.[15] The results offer the relationship between grade repetition changes and dropout rates independent of school-time fixed-effects and county trends in retention rates for the pre-policy period, as well as the effects after policy change in the school cycle 2012-2013. If the estimate of α_2 is confounded with unobserved factors that vary directly with repetition rates, for any year before policy change, α_2 would be different from zero.

Note that the coefficients in the pre-policy period are non- significantly different from zero at the 95% of confidence for all periods previous to policy reform, ruling out a possible anticipation to policy change and implying that the average coefficient α_2 represents the effect of the change in promotion rules in Mexican primary schools, as unobserved factors independent of time nor school-fixed or county changing characteristics causing varying grade repetition rates are plausibly not correlated with our estimator of interest. This offers support to the conclusion that exogenously abolishing retention in-grade caused a reduction in dropout rates in Mexico, and that the effect remains (and is statistically higher) two years after policy change.

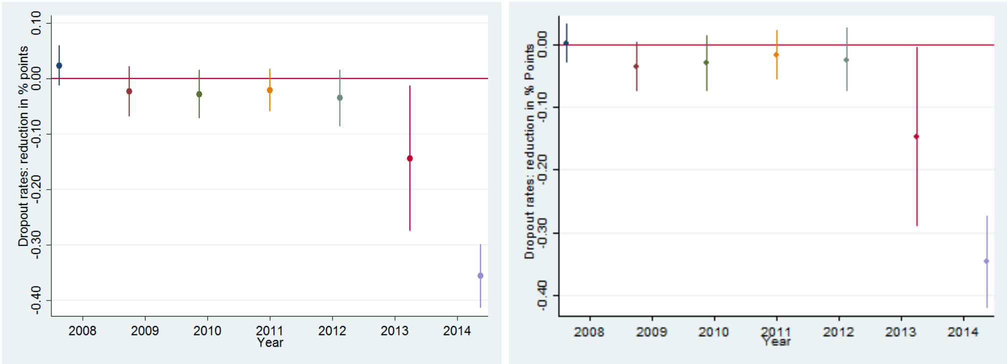
3.4.2 Robustness checks

This section presents the estimates of two different specifications to measure the effects of the reform in promotion rules. The first estimation is similar in spirit to a triple difference strategy as defined by the equation:

$$Dropout_{st} = \alpha_1 + \alpha_2 Repetition_s * Post * nonMX + Repetition_s * D_t + nonMX * D_t + D_t + D_s + \alpha_3 X_{st} + \varepsilon_{st} \quad (3.2)$$

[15]Detailed results for these specifications and another one including state-year fixed-effects are included in the Annex Table C2

FIGURE 3.6: Falsification test of a reduction in grade repetition on dropout rates in the pre- and post-policy period including school and time fixed effects (left) and school, time and county-year fixed effects (right)



Point estimates and the 95% confidence intervals are obtained from a regression including interactions between grade repetition and year dummies taking the value of 0 in the pre-policy period and 1 in the subsequent years, 2008-2012 as if policy had changed before 2012. We use 2007 repetition rates as a base. Estimations include time, school fixed-effects (left) and a combination of county-year fixed-effects (right). Regressions also include a set of school and teacher's characteristics as controls.

In this regression, α_2 gives the casual estimate of the effect of policy change in all states different from Mexico City (non-MX) on dropout rates. $Repetition_s$, Post, D_t, D_s and X_{st} are defined as before. Non-CDMX is an indicator that takes value 1 for any state different from Mexico City. Mexico City schools are used because repetition rates there did not chance dramatically after policy intervention, plausibly because they were already at a very low level, nonetheless, as the rest of Mexico they also depict an absolute downwards trend years before policy intervention as shown in Annex Figure C1. The interactions $Repetition_s * D_t$ between average repetition rates before policy change and time dummies, denote time trends that account for any time varying differences in retention rates between Mexico City's schools and the rest of Mexico apart from the change in promotion rules. Moreover there can be differences in the education conditions and policies between non-MX states and Mexico City, this is controlled by the varying state effects $non - MX * D_t$. This strategy allows schools with varying high and low repetition rates to have different time specific trends. It rests on the assumption that difference in retention trends between high and low repetition schools is identical across non-MX states and Mexico City. This is likely to hold once state specific factors are controlled for.

The second alternative estimation measures the effect of policy change using a specification similar to a triple difference but using intra-school variation as the identification strategy. As denoted before, change in promotion rules only affected 1st to 3rd grade students while the same guidelines were left for children in grades 4th to 6th. Annex Figure C2 shows first to third grade average grade repetition trends in comparison to fourth to sixth grades, which should not be affected by policy change. The lines show a reduction in repetition for lower grades and a less clear reduction in the after policy period for higher grades. The difference between the two rates in each school is used according to what is defined in the equation:

$$\sum_{g=1}^{3} Dropout_{st} - \sum_{g=4}^{6} Dropout_{st} = \alpha_1 + \alpha_2 \sum_{g=1}^{3} Repetition_s * After + D_t + D_s + \alpha_3 X_{st} + \varepsilon_{st} \tag{3.3}$$

Equation 3 takes the first differences between dropout rates from 1st to 3rd grades and 4th to 6th grades in each school S in time t=2007-2014 and uses as dependent variable, this way, all unobserved factors at the school level describing varying repetition trends between schools are differentiated out. This strategy uses 4th to 6th grade year trends as a control groups. $Repetition_s$, Post, D_t, D_s and X_{st} remain as before. Similarly, α_2 gives the casual estimate of the effect of policy change accounting for time and school fixed-effects.

The results of both specifications on dropout rates are presented in Table 3.4. They show a slightly higher average reduction of about 0.35 % points on dropout rates compared to our main estimations. Notheless the results are similar an in-line with what was already concluded before.

TABLE 3.4: Alternative models of dropout rates: Triple differences using Mexico City as control group and using differences between 1st-3rd and 4th-6th dropout rates as a dependant variable.

	(Dropout) (1)	(Dropout (2)
A. Using Mexico City as a Control		
Repetition * After * nonMX	−0·346*	
	(0·186)	
B. Differences between grades as dependent		
Repetition * After		−0·344***
		(0·036)
R-squared	0·233	0·234
Number of schools	462495	462719
School Fixed- Effects	Yes	Yes
Time Fixed-Effects	Yes	Yes
Other Controls	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, clustered on county, in parenthesis. Other controls included are teacher's education measured as the proportion of these with any given educational level in each school and "Carrera Magisterial" refering to the proportion of the 'best' teachers receiving incentives in each school. Teachers/Stud Ratio measured as the number of teachers per 100 students at the school level along with number of students is also included.

3.5 Heterogenous effects of abolishing grade retention

3.5.1 Do policy effects differ for children dropping out during the academic year or between courses?

In order to explore some underlying factors behind the effects of abolishing retention in grade, Table 3.5 includes the average results separated by intra (during the academic year) and inter-droput rates (between academic years). The estimations suggest that policy change effects come from students promoted to the next grade, who are now returning for the next academic cycle, as dropout rates between courses (inter) are reducing in a similar proportion to total dropout rates. This effect plausibly relates to students not feeling 'discouraged' after they are not failed. On the other hand, intra-course dropout rates do not seem to be affected.[16] Plausibly, children abandoning school during the

[16] In recent years, intra and inter-course dropout rates are similar on average as it can be noticed in the Annex Table C1. For wich half of the proportion of children droping out of school was not affected by the policy.

academic year are doing it for reasons less related to retention laws or school characteristics. In this regard, another variable shown in Table 3.5 related to the 'academic environment' of schools such as the proportion of teachers in "Carrera Magisterial", an incentive program directed to "the best" of them according to a set of standardized evaluations, do not seem to be related to intra-course dropout rates either, also suggesting that these type of dropout relates less to school/teacher policies and characteristics and more to other context and children/family unobserved characteristics.

TABLE 3.5: Two-way fixed-effects models of dropout rates on grade repetition

	Total (1)	Intra-course (2)	Inter-course (3)
Repetition * After	−0·254***	−0·000	−0·248***
	(0·043)	(0·019)	(0·046)
# Students	0·190***	0·068***	0·122***
	(0·008)	(0·006)	(0·005)
Students-sq.	−0·000***	−0·000***	−0·000***
	(0·000)	(0·000)	(0·000)
Teacher/Stud Ratio	−1·556***	−0·548***	−0·940***
	(0·095)	(0·050)	(0·088)
Vocational	−0·004	0·000	−0·003
	(0·005)	(0·003)	(0·004)
Bachelors	−0·008	−0·001	−0·005
	(0·005)	(0·002)	(0·004)
Postgrad	−0·013**	−0·002	−0·008*
	(0·006)	(0·003)	(0·005)
Carrera Magisterial	−0·040**	0·003	−0·047***
	(0·019)	(0·010)	(0·016)
School Fixed- Effects	Yes	Yes	Yes
Time Fixed-Effects	Yes	Yes	Yes
County * Year Fixed-Effects	Yes	Yes	Yes
R-squared	0·268	0·253	0·247
Number of schools	462798	462617	462495

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, clustered on states, in parenthesis. Teacher's education is measured as the proportion of these with any given educational level in each school. "Carrera Magisterial" is a national program that offers monetary incentives to "the best" teachers according to a set of standardized evaluations; the variable included in the model refers to the proportion of teachers receiving such incentives in each school. Teachers/Stud Ratio is measured as the number of teachers per 100 students at the school level.

3.5.2 Do policy effects differ by school's socioeconomic contexts?

One important test for the results is whether the effects of abolishing retention in-grade hold for different socioeconomic contexts or not. For example, having no impacts of policy change in poorer schools may undone the whole purpose of a policy seeking to help vulnerable kids to stay longer in school. In this regard, Table 3.6 presents the estimation defined in Equation 1 using the same sample of schools separated by low- and high-SES including school, time and county-year fixed effects.[17] The results remain similar to those shown in Table 3.3 as there is a significant dropout rates reduction in poorer and richer schools and the inter-course dropout rates remain driving general results in both type of schools. This is confirmed by the estimations presented in Table 3.7 which separates schools above (poorer) and below (richer) each state's average of students with a PROSPERA scholarship.

TABLE 3.6: Heterogenous effects between low- and mid/high-SES schools

	Low-SES			Mid to High-SES		
	Total (1)	Intra-course (2)	Inter-course (3)	Total (4)	Intra-course (5)	Inter-course (6)
Repetition*After	−0·206***	−0·009	−0·212***	−0·294***	0·011	−0·281***
	(0·060)	(0·017)	(0·066)	(0·055)	(0·030)	(0·053)
School Fixed- Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
County * Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0·268	0·243	0·250	0·273	0·264	0·250
Number of schools	161094	161009	160957	301704	301608	301538

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, clustered on states, in parenthesis. Point estimates come from the interaction between school's grade repetition rates and a dummy taking the value of 1 in the after-policy period. Estimations include time and school fixed-effects and a set of school and teacher's characteristics as controls as described in equation 1.

Note that, when comparing the estimations on total and inter-course dropout rates in both sets of results presented in Tables 3.6 and 3.7, the effect of policy change is not significantly different in richer schools once standard errors are considered. For example, the effect observed in Low-SES schools on total dropout rates is of -0.21 % points (S.E. of 0.06) vs -0.29 % points (S.E. of 0.06). This same conclusion holds for the inter-course dropout rates.

[17]The remaining estimations will also include this setup as according to the Falsification tests presented above, seems to control for pre-trends in a more robust way

TABLE 3.7: Heterogenous effects: Proportion of students receiving a PROGRESA scholarship

	Schools Above State Average			Schools Below State Average		
	Total (1)	Intra-course (2)	Inter-course (3)	Total (4)	Intra-course (5)	Inter-course (6)
Repetition * After	−0·198***	−0·011	−0·201***	−0·293***	0·009	−0·276***
	(0·045)	(0·018)	(0·043)	(0·058)	(0·028)	(0·062)
School Fixed- Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
County * Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0·277	0·254	0·263	0·312	0·308	0·287
Number of schools	175407	175347	175299	281587	281465	281385

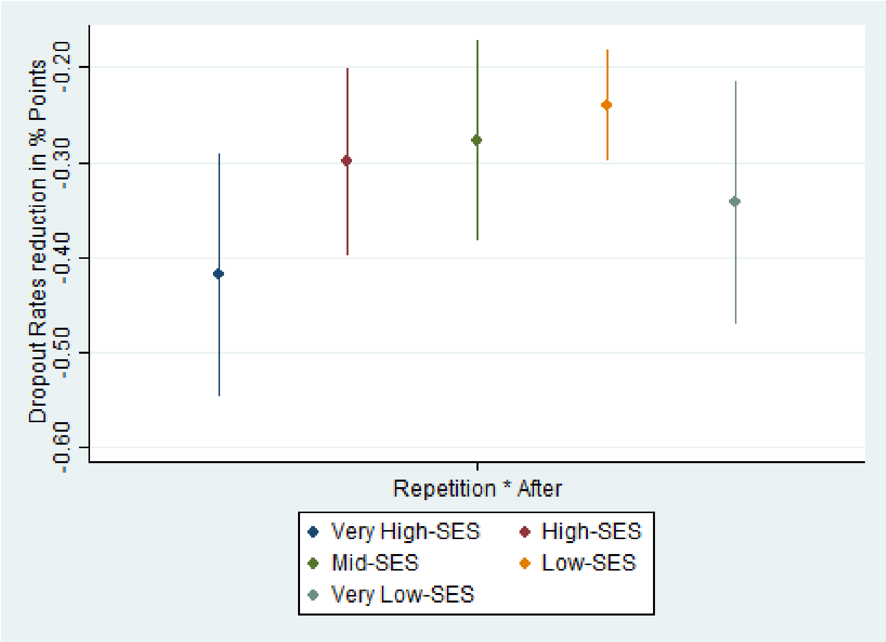
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, clustered on state, in parenthesis. Point estimates come from the interaction between school's grade repetition rates and a dummy taking the value of 1 in the after-policy period. Estimations include time and school fixed-effects and a set of school and teacher's characteristics as controls as described in Equation 1.

Further results by SES are presented in Figure 3.7 which separates schools by their socioeconomic status as reported in the ENLACE dataset. The results allow to confirm no significant differences for most schools with different levels of privation. Although, results suggest a statistically significant higher effect for the Very High-SES compared to Low-SES schools. This effect is sensible as one would expect that dropout rates for poorer children would be less driven by school promotion policies and more by their context. Additionally, a possibility behind higher effects in richer schools may relate to the type of dropout observed in such contexts, as plausibly, richer children may temporarily dropout to return later to finish their studies and now they do not find the need to do it. Contrary, effects in poorer schools may be lower because dropouts in such contexts may be leaving schools definitively.

Another differentiated effect of interest to test varying effects across different socioeconomic backgrounds is the one between public and private schools as both differ in the composition of their students and their previous trends of grade repetition as presented before in Figure 3.5. Table 3.8 shows the results by type of school. First thing to note is that points estimates are more similar between these two types of schools compared to those observed between low- and high-SES schools. This understandable as 25% of public schools are also located in high-SES localities, although this is still a small proportion compared to the 80% of private schools. The estimations exhibit that the effect of policy change in total dropout rates is fairly similar in public and private institutions, i.e. 0.25 % points (S.E of 0.02) and 0.23 % points (S.E.of 0.07), respectively. Similar

FIGURE 3.7: Policy effects on dropout rates by different categories of SES.



Point estimates and 95% confidence intervals come from the interaction between grade repetition rates in each school in the pre- and post-policy period and a dummy taking the value of 1 in the after policy period and zero otherwise. The estimation includes time and school fixed-effects and a set of school and teacher's characteristics as controls

point estimators should not be surprising as average repetition rates in the pre-policy period are higher in public institutions than in private, and thus one can expect a higher relative impact in public schools.

TABLE 3.8: Heterogenous effects between private and public schools

	Public			Private		
	Total (1)	Intra-course (2)	Inter-course (3)	Total (4)	Intra-course (5)	Inter-course (6)
Repetition*After	$-0{\cdot}254^{***}$	$-0{\cdot}008$	$-0{\cdot}249^{***}$	$-0{\cdot}230^{**}$	$0{\cdot}054$	$-0{\cdot}242^{***}$
	(0·043)	(0·020)	(0·046)	(0·099)	(0·059)	(0·087)
School Fixed- Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
County * Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0·263	0·262	0·229	0·309	0·243	0·321
Number of Schools	412639	412511	412414	50159	50106	50081

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, clustered on state, in parenthesis. Point estimates come from the interaction between school's grade repetition rates and a dummy taking the value of 1 in the after-policy period. Estimations include time and school fixed-effects and a set of school and teacher's characteristics as controls as described in Equation 1.

3.5.3 Do family investments modify the effect of abolishing retention in-grade?

A possibly different way to observe relevant differences between varying socioeconomic contexts is to asses the effect of the policy change conditional on family investments, as these may undo the intended effect of abolishing retention in-grade for some disadvantaged groups. If for example, abolishing grade retention does not have an effect on children who receive less pecuniary and non-pecuniary investments from their parents relative to their relative socioeconomic context, the whole purpose of the policy, assuming that children receiving lower family investments are the most likely to abandon school, can be undone.

As discussed, the data at hand does not offer individual information on children's background. Thus, we use the proportion of children in any given school with one to three years of pre-school education as proxy for average family investments. Specifically, if a school has a proportion of children with one year of preschool that is higher than the average of the State where the school is located, it takes the value of 1 and zero otherwise, and so forth for schools with proportions above averages of children with two and three years of kindergarten. The intuition is that schools with higher proportions of children who received more years of preschool concentrate pupils who receive more investments from their parents and the opposite in the case of those schools below the average.

In this regard, Table 3.9 shows the separated effects in schools with lower and higher proportions of children with pre-school education In panel A we can observe point estimators for public schools and in Panel B for private schools. We separate the results by Public and Private schools as themselves represent different investments on children. Column 1 presents the effect of policy change on schools with a proportion of children with at least one year of pre-school education below their state's average. Column 2 to 4 presents the results for schools above the average proportion of children with one to three years of pre-school.

Results show, after considering the standard errors of each estimation, that in public schools there are no significant differences between schools with lower or higher proportions of children with pre-school education. Regarding private schools, the higher standard errors possibly coming from a reduction of the number of observations, do not allow to confirm significant effects in columns 3 and 4. This is sensible as this would indicate that more advantaged children who receive the highest investments combining more years of pre-school and education coming from private schools are not presenting effects of the policy change, plausibly because grade repetition rates are close to zero in such contexts. Similarly, there seems to be a higher effect of the policy on schools with a higher proportion of children with 1 year of preschool (-0.62 % points, S.E. of 0.31) versus schools below the average proportion as reported in Column 1 (-0.21 % Points, S.E. of 0.07) . This suggests that the policy is presenting lower effects in private schools that concentrate children who receive relatively less parental investments.

TABLE 3.9: Policy effects on public and private schools with higher proportions of children with pre-school education

Preschool Education	Below Average (1)	1 Year (2)	2 years (3)	3 years (4)
A. Public Schools				
Repeat*After	−0·236***	−0·243***	−0·262***	−0·281***
	(0·053)	(0·059)	(0·046)	(0·057)
R-squared	0·304	0·405	0·428	0·406
Number of schools	280057	120159	74808	95390
B. Private Schools				
Repeat*After	−0·206***	−0·619*	−0·417	−0·706
	(0·072)	(0·305)	(0·271)	(0·444)
R-squared	0·324	0·524	0·537	0·515
Number of schools	42548	6096	4106	5196
School Fixed- Effects	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes
County*Year F.E.	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, clustered on state, in parenthesis. Point estimates come from the interaction between school's grade repetition rates and a dummy taking the value of 1 in the after-policy period. Estimations include a set of fixed-effects and controls as described in Equation 1.

3.5.4 Does abolishing retention in grade affect pupil's performance?

One of the most common arguments in favor of retention in grade, is that it acts as a deterrent to poor school performance by inflicting a high penalty on under-performers, working as an incentive for students to increase their efforts. Therefore a plausible unintended outcome coming from the automatic promotion of students, is a possible negative effect of eliminating the "threat" of grade repetition on children's performance at school. It is also worth considering that the effect of abolishing grade repetition on performance, may vary by socioeconomic status and may be more detrimental in the case of children from more deprived contexts or with less motivated parents.

In order to explore the consequence of changing promotion guidelines in Mexico, we take advantage of the same set-up presented in Equation (1) to evaluate the effects on standardized Math and Spanish test scores. We use the ENLACE results for that purpose. Dismally, the analysis is constrained by the fact that the test was only applied until the school year 2012-2013 and that it is only applied to children in 3th to 6th grade. Conveniently, we can study the effects of policy change on test scores of 3rd graders as change in promotion guidelines started from September 2012, whereas the last round of the ENLACE test was applied at the end of the school year, in June 2013. So it is important to highlight that the effect we are exploring here is not that of the "automatic promotion" as we would need to observe test results of children who shouldn't have been promoted in the absence of the policy, consequently, we study the elimination of the "threat" of grade retention in the year were the test score was applied, as this may discourage children to perform.

Table 3.10 shows the effects of policy change on mathematics in Panel A and on Spanish in Panel B. The columns show different heterogeneous results for low- and high-SES schools, both public and private. The estimations in Column 1 suggest a statistically significant but economically insignificant positive effect of abolishing grade retention in both Mathematics and Spanish test scores of 0.4-0.7% of a Standard Deviation. These point estimators translate into a marginal gain of just 4 to 7 points in a test with an average of 500 points. Note that when the schools are separated into low and high-SES, the effect becomes non-different from zero. In the overall results in Table 3.10 suggest that eliminating the "threat" of grade repetition could have not reduced the effort of third graders. An alternative interpretation is that policy change could have increased school resources and efforts devoted to children who progress but would not have progressed in the absence on the reform. Unfortunately, this is a question that remains open given the data at hand.

TABLE 3.10: Two-way fixed-effects models of standardized math and Spanish test scores on grade repetition

	Average	Low-SES		High-SES	
		Public	Private	Public	Private
	(1)	(2)	(3)	(4)	(5)
A. Mathematics					
Repetition*After	0·004**	−0·002	0·007	0·005	−0·002
	(0·002)	(0·004)	(0·013)	(0·003)	(0·001)
R-squared	0·547	0·467	0·850	0·502	0·863
Number of schools	370904	112669	2968	218965	36302
B. Spanish					
Repetition*After	0·007***	−0·000	0·018	0·007**	0·000
	(0·002)	(0·004)	(0·018)	(0·003)	(0·001)
R-squared	0·605	0·462	0·902	0·554	0·910
Number of schools	370844	112633	2968	218952	36291
School Fixed- Effects	Yes	Yes	Yes	Yes	Yes
Time Fixed-Effects	Yes	Yes	Yes	Yes	Yes
County*Year F.E.	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, clustered on state, in parenthesis. Regressions include a full set of controls such as teacher's average education at the school level, if they participate in "Carrera Magisterial", number of students and teacher-to-student ratio, as well as the number of student's tested and the proportion of results considered as unreliable in each school.

3.6 Conclusions

This paper evaluated the impact effects of an exogenous policy change that took place in 2012 which modified Mexico's grade retention laws for 1st to 3rd grade students on dropout rates. The data at hand offered the possibility to study a panel of schools between 2007 to 2014 and to explore different heterogeneous effects by socioeconomic status and type of schools. Additionally, it was possible to summarily test if abolishing grade repetition 'threat' would affect children's performance.

The main results obtained exhibit a significant effect of abolishing grade retention on dropout rates of approximately 0.25% points. General findings do not seem to respond to pre-trends in retention rates as placebo tests probe that our identification strategy manages to isolate the effect of the exogenous change in retention laws. Similarly, alternative specifications allowed to confirm similar effects.

The average reduction in dropout rates is not minimal if we consider that the average before policy change was already close to a 1%. Such effects are observed one and two years after policy change. Indeed this is a very short span of time and we acknowledge that they are rather preliminary and will be more relevant when more data on dropout and pupil's performance is available as it will be interesting to analyze if the short term effects remain once children grow older and go to secondary school where recent dropout rates in Mexico can be as high as 8%. Nonetheless, these results are a good a first approach to the possible effects of a policy that as mentioned, it has now been revoked, despite the lack of an impact evaluation on relevant outcomes.

Furthermore, results show that the effects of policy change are rather stable across different socioeconomic backgrounds and in private and public schools and it does not seem to affect children's performance. This offers support to the idea that grade repetition is not only dependent of children's socioeconomic and cultural conditions and that education policies have an important role. In fact, we have shown that a low-cost policy which only needed to modify retention laws independent of other inputs at the school level, can directly enhance children's overall achievement without seemingly affecting pupil's general performance. This is of great relevance in the Mexican context where authorities face an important challenge to improve education outcomes.

General Conclusions

This thesis focuses on the evaluation of three interventions at the school and student level in Mexico with the use of quasi-experimental techniques. The first chapter analysed the effect of a school breakfast program (SBP) that as been active for decades. Second chapter focused on exploring the impacts of one of the most important education policies implemented in Mexico in the last ten years, the Full-Time Schools Program or PETC. Last chapter discussed the effects of abolishing grade retention for the first grades in primary schools.

The main results regarding the SBP show that program's targeting is a central issue. Evidence shows that the treatment does not depend on children's characteristics like height and weight, as it is claimed by national education authorities. Contrary, results suggest an administrative bias in the allocation of breakfasts, plausibly subjected to characteristics of schools. For example, results suggest that schools with more students and with more prepared principals have a higher chance to participate of the breakfast policy, when these are usually the most deprived. The estimations also show a consistent effect on children's weight which may push them above their standardise average by age and sex. Such effects could be even more detrimental on children at the bottom half of the income distribution. Results on other outcomes such as cognitive skills and grade repetition are not significant.

The evaluation of the SBP offers some interesting conclusions regarding the future of the policy. It opens the debate on the objectives and targeting of the program in a context where nutritional challenges are different. When the SBP started, its goal was to reduce children's under-nutrition, but nowadays Mexican children are in a very different situation, facing now a greater risk of obesity rather than low-weight. Nonetheless, obvious limitations constrain the strength of the conclusions reached in Chapter 1. The sample used is rather small, and although it comes from a nationally representative survey, sample size most surely affects the efficiency of our estimations and some results may be significant with a different and/or larger sample. This opens a space for future research. Possibly an experimental set-up would be a suitable alternative to test different nutrition interventions directed to improve how much and what children eat at schools to set the initial stages of an updated children's nutrition program.

The empirical research presented in this document also included an evaluation to the school day reform, the flagship education policy of the previous federal administration (2006-2012) and one of the most important in the present government. The program implied a big investment on material and teaching resources and nowadays it continues expanding at a fast pace. The evaluation presented its effects on pupil's achievement between 2007 and 2013 using standardised test scores of mathematics and Spanish as an outcome. The results showed positive impacts of up to 0.11 SD after 4 years of treatment on both subjects. The results also show an impact almost three times higher in poorer schools. Estimations also suggest that for the case of mathematics, after four years of treatment there is a clear reduction in the proportion of students graded as insufficient and an increase of those with excellent results. These results are of key relevance to argue that less skilled children even in deprived environments, are also benefiting for longer school days.

These conclusions are helpful to guide the continuous fast expansion of the program which aims to grow from around 6,000 schools treated in 2012 to 40,000 by 2018. Resources may be more efficiently invested if they concentrate in poorer schools in accordance to the empirical evidence reported above. Nevertheless, it is important to mention that, although the results presented in Chapter 2 are robust to different specifications, the program has grown quickly and its effectiveness has surely changed, for which is important to update the evaluation of its impacts and make sure that with a newer, bigger and more challenging sample in terms of policy control and intervention, such results remain positive.

The final chapter evaluates the arguably causal impacts of a change to the general rules for the student's evaluation, promotion and certification in basic education. The reform only affected children from first to third grades. The main results obtained exhibit a significant effect of abolishing grade retention on dropout rates of approximately 0.25% points with no seeming effects of eliminating the threat of grade repetition on student's test score. The general findings are robust to different specifications and falsification tests and are consistent across different socioeconomic backgrounds, types of schools and a proxy for family investments. Unfortunately, this policy was modified again only one year later in order to bring back the older laws allowing grade retention of 2nd and

3rd graders despite the lack of any empirical evidence at that moment on the Mexican case.

The estimations are informative of the short term effects of the program and give some support to the continuation and possibly, the expansion of 'social promotion' to other grades in the Mexican education system rather than continuing with a punitive system, specially in secondary schools where dropout rates can be as much as seven times higher than in primary schools. Unfortunately, possible experimental setups to evaluate such expansions are limited due to the fact that changes in promotion laws in Mexico have to happen at the federal level. Nevertheless, the evidence presented in Chapter 3 is a good start to open the debate and develop future research on the topic as the question on the mid-run effects of abolishing grade repetition is still open considering that the short term effects on reducing dropouts can still vanish once children go to higher levels of education.

In the overall, this thesis sought to contribute to the reduced ambit of the impact evaluation of education policies in Mexico. The main objective was to discuss, as much as possible, the causal effects of specific policies, but with the broader intention to point out the urgent need to generate or redesign interventions based on a deeper understanding of their overall achievements. This is important not only because education policies make use of a large amount of public resources, but also because the mistakes that governments make in this respect can potentially affect entire generations and may have lasting detrimental effects, specially, among the most vulnerable.

Annex tables and figures: Chapter 1

TABLE A1: Determinants of the assignment into treatment SBP in Mexico. BMI, Height(Cm.), Weight(Kg.) and Height over Weight by age and sex

	(1) SBP	(2) SBP	(3) SBP	(4) SBP	(5) SBP
Body Mass Index	-0.007 (0.033)	0.007 (0.034)	0.044 (0.036)	0.069 (0.059)	0.090 (0.061)
Height(Cm.)	-0.006 (0.017)	0.003 (0.017)	0.011 (0.018)	-0.014 (0.039)	-0.006 (0.037)
Weight(Kg.)	-0.004 (0.020)	-0.001 (0.020)	0.015 (0.021)	0.039 (0.039)	0.046 (0.040)
Height/Weight	-0.060 (0.089)	-0.006 (0.093)	0.120 (0.101)	0.119 (0.159)	0.198 (0.162)
Observations	587	587	587	317	317
Kids Controls	X	X	X	X	X
Context Controls		X	X		X
Progresa		X	X		X
School Controls			X	X	X
Municipality Fixed-effects				X	X

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A2: Treated and untreated average of covariates used in the PSM for the panel of kids 2002-2005: Nearest Neighbour (5)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.796	9.922	-0.126
Log of Income Sqrd.	96.947	99.221	-2.274
Mother Schooling	7.043	7.030	0.014
Father Schooling	7.450	7.543	-0.092
Siblings	2.485	2.359	0.126
PROGRESA	0.199	0.145	0.054
Girl	0.515	0.463	0.052
Age	8.212	8.123	0.089
IQ	0.052	-0.002	0.055
Standard. Height	0.019	0.064	-0.046
Standard. Weight	-0.039	0.041	-0.080
Illness dummy	0.610	0.609	0.001
Multi-grade School	0.130	0.135	-0.005
#Teachers	10.377	10.193	0.184
#Teachers Sqrd.	127.790	124.980	2.810
#Students	313.350	289.830	23.520
Principal Education	1.745	1.607	0.138
Principal Female	0.225	0.331	-0.106
Principal Another Job	0.420	0.373	0.047
Cleaning Personnel	0.658	0.707	-0.049
Bathroom/Water/Soup	0.429	0.376	0.052
Good Ceiling Condition	0.900	0.884	0.016
Region	3.597	3.485	0.112

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A3: Treated and untreated average of the covariates used in the PSM for the Panel of Kids 2002-2005: Mahalanobis Distance (5)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.7859	9.9333	-0.1474
Log of Income Sqrd.	96.752	99.337	-2.585
Mother Schooling	6.9707	7.154	-0.1833
Father Schooling	7.3305	7.8259	-0.4954
Siblings	2.5607	2.5598	0.0009*
PROGRESA	0.21339	0.12469	0.0887
Girl	0.5272	0.50795	0.01925
Age	8.2218	8.1498	0.072
IQ	0.04977	-0.04482	0.09459
Strd. Height	0.01892	-0.03698	0.0559
Strd. Weight	-0.03185	-0.16	0.12815
Illness dummy	0.61925	0.64017	-0.02092
Multi-grade School	0.12552	0.06695	0.05857
#Teachers	10.331	9.9113	0.4197**
#Teachers Sqrd.	126.22	112.26	13.96**
#Students	310.63	300.33	10.3
Principal Education	1.7866	1.5456	0.241
Principal Female	0.21757	0.18912	0.02845
Principal Another Job	0.43933	0.34895	0.09038
Cleaning Personnel	0.63598	0.77741	-0.14143
Bathroom/Water/Soup	0.41423	0.21423	0.2
Good Ceiling Condition	0.90377	0.89205	0.01172
Region	3.6778	3.3582	0.3196*

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A4: Treated and untreated average of the covariates used in the PSM for the panel of kids 2002-2005: Radius with caliper (0.001)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.8948	9.9656	-0.0708
Log of Income Sqrd.	99.09	100.16	-1.07
Mother Schooling	7.3038	6.9251	0.3787
Father Schooling	7.4557	7.5646	-0.1089
Siblings	2.5823	2.4622	0.1201
PROGRESA	0.18987	0.18776	0.00211
Girl	0.5443	0.47996	0.06434
Age	8.2532	8.1846	0.0686
IQ	-0.0289	0.0801	-0.109
Std. Height	-0.04471	-0.03074	-0.01397
Std. Weight	0.08698	-0.05193	0.13891
Illness dummy	0.58228	0.56624	0.01604
Multi-grade School	0.12658	0.09283	0.03375
#Teachers	10.152	9.9681	0.1839
#Teachers Sqrd.	123.47	119.68	3.79
#Students	297.24	282.45	14.79
Principal Education	1.557	1.6589	-0.1019
Principal Female	0.26582	0.26962	-0.0038
Principal Another Job	0.44304	0.45844	-0.0154
Cleaning Personnel	0.73418	0.73312	0.00106
Bathroom/Water/Soup	0.34177	0.34304	-0.00127
Good Ceiling Condition	0.87342	0.84705	0.02637
Region	3.5823	3.289	0.2933

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A5: Treated and untreated average of the covariates used in the PSM for the panel of kids 2005-2005: Radius with caliper (0.005)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.836	9.877	-0.042
Log of Income Sqrd.	97.762	98.276	-0.514
Mother Schooling	7.012	6.972	0.040
Father Schooling	7.406	7.690	-0.284
Siblings	2.558	2.553	0.005
PROGRESA	0.176	0.155	0.021
Girl	0.503	0.520	-0.017
Age	8.164	8.181	-0.017
IQ	-0.027	0.140	-0.167
Std. Height	0.024	0.049	-0.025
Std. Weight	0.019	-0.007	0.026
Illness dummy	0.612	0.569	0.043
Multi-grade School	0.121	0.186	-0.065
#Teachers	10.206	9.799	0.407
#Teachers Sqrd.	126.420	117.070	9.350
#Students	301.410	287.880	13.530
Principal Education	1.588	1.570	0.018
Principal Female	0.261	0.285	-0.024
Principal Another Job	0.352	0.392	-0.041
Cleaning Personnel	0.673	0.673	-0.001
Bathroom/Water/Soup	0.352	0.367	-0.015
Good Ceiling Condition	0.879	0.888	-0.010
Region	3.491	3.515	-0.024

$* p < 0.1$, $** p < 0.05$, $*** p < 0.01$

TABLE A6: Treated and untreated average of the covariates used in the PSM for the panel of kids 2002-2005: Kernel bandwidth (0.001)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.8948	9.9783	-0.084
Log of Income Sqrd.	99.09	100.38	-1.290
Mother Schooling	7.3038	6.8618	0.442
Father Schooling	7.4557	7.5488	-0.093
Siblings	2.5823	2.4575	0.125
PROGRESA	0.18987	0.18714	0.003
Girl	0.5443	0.47997	0.064
Age	8.2532	8.1669	0.086
IQ	-0.0289	0.07544	-0.104
Std. Height	-0.04471	-0.01497	-0.030
Std. Weight	0.08698	-0.03024	0.117
Illness dummy	0.58228	0.5554	0.027
Multi-grade School	0.12658	0.10013	0.026
#Teachers	10.152	10.025	0.127
#Teachers Sqrd.	123.47	121.56	1.910
#Students	297.24	284.69	12.550
Principal Education	1.557	1.6547	-0.098
Principal Female	0.26582	0.28698	-0.021
Principal Another Job	0.44304	0.46687	-0.024
Cleaning Personnel	0.73418	0.73968	-0.005
Bathroom/Water/Soup	0.34177	0.34658	-0.005
Good Ceiling Condition	0.87342	0.85057	0.023
Region	3.5823	3.315	0.267

$* p < 0.1$, $** p < 0.05$, $*** p < 0.01$

TABLE A7: Treated and untreated average of the covariates used in the PSM for the panel of kids 2002-2005: Kernel bandwidth (0.005)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.8356	9.8992	-0.064
Log of Income Sqrd.	97.762	98.713	-0.951
Mother Schooling	7.0121	7.0528	-0.041
Father Schooling	7.4061	7.7142	-0.308
Siblings	2.5576	2.5531	0.004
PROGRESA	0.17576	0.14583	0.030
Girl	0.50303	0.50874	-0.006
Age	8.1636	8.1894	-0.026
IQ	-0.02674	0.17012	-0.197
Std. Height	0.0243	0.05846	-0.034
Std. Weight	0.01927	0.00516	0.014
Illness dummy	0.61212	0.5635	0.049
Multi-grade School	0.12121	0.1901	-0.069
#Teachers	10.206	9.8616	0.344
#Teachers Sqrd.	126.42	118.43	7.990
#Students	301.41	289.53	11.880
Principal Education	1.5879	1.5812	0.007
Principal Female	0.26061	0.2814	-0.021
Principal Another Job	0.35152	0.39558	-0.044
Cleaning Personnel	0.67273	0.68889	-0.016
Bathroom/Water/Soup	0.35152	0.36488	-0.013
Good Ceiling Condition	0.87879	0.88188	-0.003
Region	3.4909	3.5029	-0.012

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A8: Treated and untreated average of covariates used in the PSM for the cohort of kids 2002: Nearest Neighbour (5)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.8734	9.9583	-0.085
Log of Income Sqrd.	98.394	99.936	-1.542
Age	8.3699	8.4491	-0.079
Mother Schooling	6.8934	6.8488	0.045
Father Schooling	7.3981	7.2792	0.119
Siblings	2.4702	2.5037	-0.033
PROGRESA	0.19749	0.14499	0.053
Girl	0.50157	0.48685	0.015
Multi-grade School	0.15674	0.12081	0.036
#Teachers	10.345	10.255	0.090
#Teachers Sqrd.	129.08	128.01	1.070
#Students	313.5	299.84	13.660
Principal Education	1.7241	1.5894	0.135
Principal Female	0.24451	0.28743	-0.043
Principal Another Job	0.44514	0.42283	0.022
Cleaning Personnel	0.67085	0.72645	-0.056
Bathroom/Water/Soup	0.43887	0.36185	0.077
Good Ceiling Condition	0.90282	0.87943	0.023
Region	3.5549	3.5101	0.045

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A9: Treated and untreated average of the covariates used in the PSM for the Cohort of Kids 2002: Mahalanobis Distance (5)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.8475	10.025	-0.178
Log of Income Sqrd.	97.924	101.08	-3.156
Age	8.3795	8.3922	-0.013
Mother Schooling	6.8133	7.3127	-0.499
Father Schooling	7.2711	7.8946	-0.624
Siblings	2.5512	2.2922	0.259
PROGRESA	0.21988	0.11446	0.105
Girl	0.51807	0.50602	0.012
Multi-grade School	0.1506	0.06687	0.084
#Teachers	10.292	10.347	-0.055**
#Teachers Sqrd.	127.2	123.7	3.500**
#Students	310.31	320.07	-9.760
Principal Education	1.7741	1.5934	0.181
Principal Female	0.23494	0.21928	0.016
Principal Another Job	0.46687	0.38614	0.081*
Cleaning Personnel	0.64458	0.83855	-0.194
Bathroom/Water/Soup	0.42169	0.25482	0.167
Good Ceiling Condition	0.90663	0.9012	0.005
Region	3.6506	3.3367	0.314

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A10: Treated and untreated average of the covariates used in the PSM for the cohort of kids 2002: Radius with caliper (0.001)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	10.114	10.098	0.016
Log of Income Sqrd.	102.94	102.69	0.250
Age	8.3464	8.561	-0.215
Mother Schooling	7.0261	7.1389	-0.113
Father Schooling	7.3203	8.122	-0.802***
Siblings	2.5294	2.317	0.212
PROGRESA	0.14379	0.13671	0.007
Girl	0.4902	0.51198	-0.022
Multi-grade School	0.14379	0.08497	0.059**
#Teachers	10.261	10.278	-0.017
#Teachers Sqrd.	130.26	124.45	5.810
#Students	300.64	309.08	-8.440
Principal Education	1.6405	1.6122	0.028
Principal Female	0.36601	0.27832	0.088
Principal Another Job	0.45098	0.40632	0.045
Cleaning Personnel	0.75163	0.79847	-0.047
Bathroom/Water/Soup	0.34641	0.34259	0.004
Good Ceiling Condition	0.86928	0.8671	0.002
Region	3.4183	3.3154	0.103

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A11: Treated and untreated average of the covariates used in the PSM for the cohort of kids 2002: Radius with caliper (0.005)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.9372	10.03	-0.093
Log of Income Sqrd.	99.656	101.33	-1.674
Age	8.3466	8.4379	-0.091
Mother Schooling	6.9044	6.6949	0.210
Father Schooling	7.4263	7.3224	0.104
Siblings	2.4622	2.2887	0.174
PROGRESA	0.16733	0.15048	0.017
Girl	0.50199	0.50957	-0.008
Multi-grade School	0.15139	0.12394	0.027
#Teachers	10.247	10.346	-0.099
#Teachers Sqrd.	128.93	128.06	0.870
#Students	301.84	305.25	-3.410
Principal Education	1.5936	1.5756	0.018
Principal Female	0.29084	0.28141	0.009
Principal Another Job	0.40239	0.42937	-0.027
Cleaning Personnel	0.69323	0.73075	-0.038
Bathroom/Water/Soup	0.36653	0.38478	-0.018***
Good Ceiling Condition	0.88446	0.88822	-0.004
Region	3.3625	3.4264	-0.064

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A12: Treated and untreated average of the covariates used in the PSM for the cohort of kids 2002: Kernel bandwidth (0.001)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	10.114	10.086	0.028
Log of Income Sqrd.	102.94	102.44	0.500
Age	8.3464	8.5695	-0.223
Mother Schooling	7.0261	7.1341	-0.108
Father Schooling	7.3203	8.1713	-0.851***
Siblings	2.5294	2.3214	0.208
PROGRESA	0.14379	0.14106	0.003
Girl	0.4902	0.49718	-0.007
Multi-grade School	0.14379	0.08783	0.056**
#Teachers	10.261	10.299	-0.038
#Teachers Sqrd.	130.26	124.86	5.400
#Students	300.64	309.56	-8.920
Principal Education	1.6405	1.6004	0.040
Principal Female	0.36601	0.28357	0.082
Principal Another Job	0.45098	0.39698	0.054
Cleaning Personnel	0.75163	0.7978	-0.046
Bathroom/Water/Soup	0.34641	0.33981	0.007
Good Ceiling Condition	0.86928	0.862	0.007
Region	3.4183	3.3346	0.084

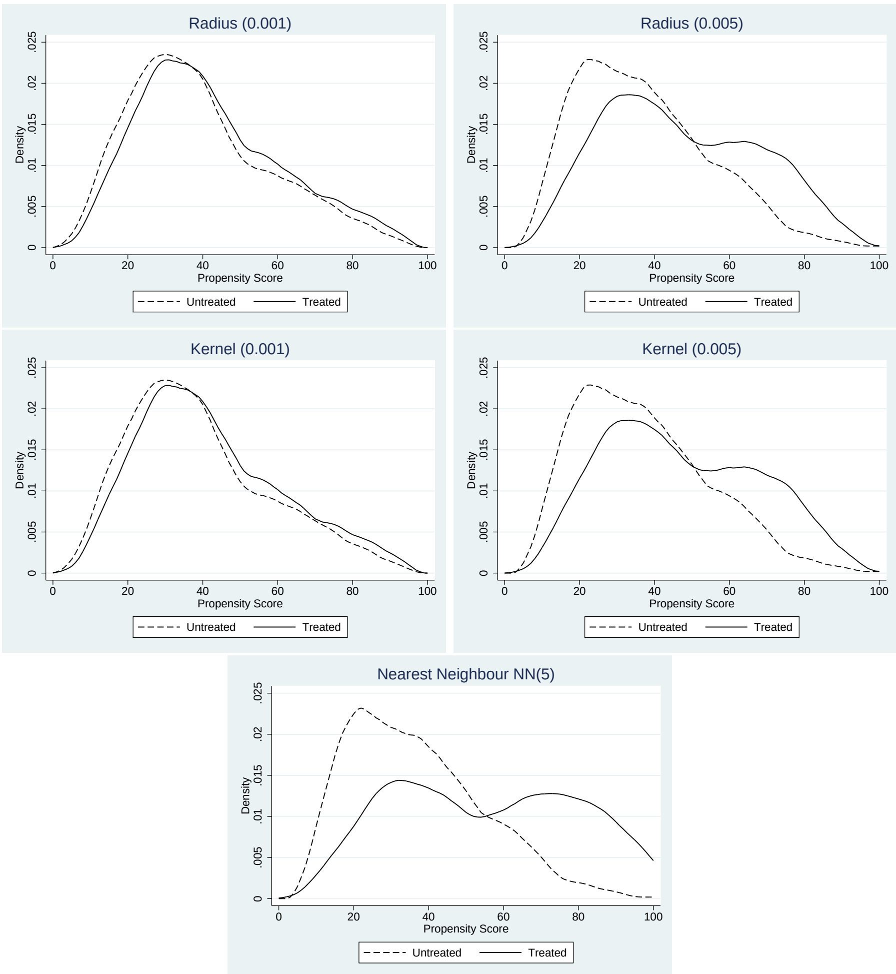
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A13: Treated and untreated average of the covariates used in the PSM for the cohort of kids 2002: Kernel bandwidth (0.005)

Variable	*Treated Mean*	*Control Mean*	*Difference*
Log of income	9.9372	10.036	-0.099
Log of Income Sqrd.	99.656	101.45	-1.794
Age	8.3466	8.462	-0.115
Mother Schooling	6.9044	6.7026	0.202
Father Schooling	7.4263	7.3218	0.105
Siblings	2.4622	2.2976	0.165
PROGRESA	0.16733	0.15212	0.015
Girl	0.50199	0.51285	-0.011
Multi-grade School	0.15139	0.12186	0.030
#Teachers	10.247	10.276	-0.029
#Teachers Sqrd.	128.93	126.14	2.790
#Students	301.84	303.06	-1.220
Principal Education	1.5936	1.5752	0.018
Principal Female	0.29084	0.28479	0.006
Principal Another Job	0.40239	0.41931	-0.017
Cleaning Personnel	0.69323	0.73286	-0.040***
Bathroom/Water/Soup	0.36653	0.38302	-0.016*
Good Ceiling Condition	0.88446	0.88311	0.001
Region	3.3625	3.4215	-0.059

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

FIGURE A1: Propensity Score After Matching cohort of Kids 2002



Annex tables and figures: Chapter 2

TABLE B1: Suggested Time Table for Full-Time Primary Schools in Mexico

	Monday	*Tuesday*	*Wednesday*	*Thursday*	*Friday*
8:30-12:30	**Math** **Spanish** **Spanish** Break Arts	**Math** **Spanish** **Science** Break Geography	**Math** **Spanish** **Science** Break Geography	**Math** **Spanish** **Science** Break Civism	**Math** **Spanish** History Break Sports
12:30-13:00	English	Arts	Sports	Sports	Arts
13:00-14:00	Food Break	Food Break	Food Break	Food Break	Food Break
14:00-14:15	Time out	Timeout	Timeout	Timeout	Timeout
14:15-14:45	**Tutoring***	**Tutoring**	**Tutoring**	**Tutoring**	**Tutoring**
14:45-15:15	IT	Social	IT	Social	Social
15:15-16:00	Sports	IT	English	Arts	English
16:00-17:00	Planning**	Planning	Planning	Planning	Planning

Source: Secretariat of Basic Education
**To hep students with homework and/or further instruction on core subjects*
***For the professors to plan and structure their lessons or talk to parents.*
****Arts, English and IT are new to the curricula.*

FIGURE B1: Trends of Spanish average scores ENLACE by treatment status

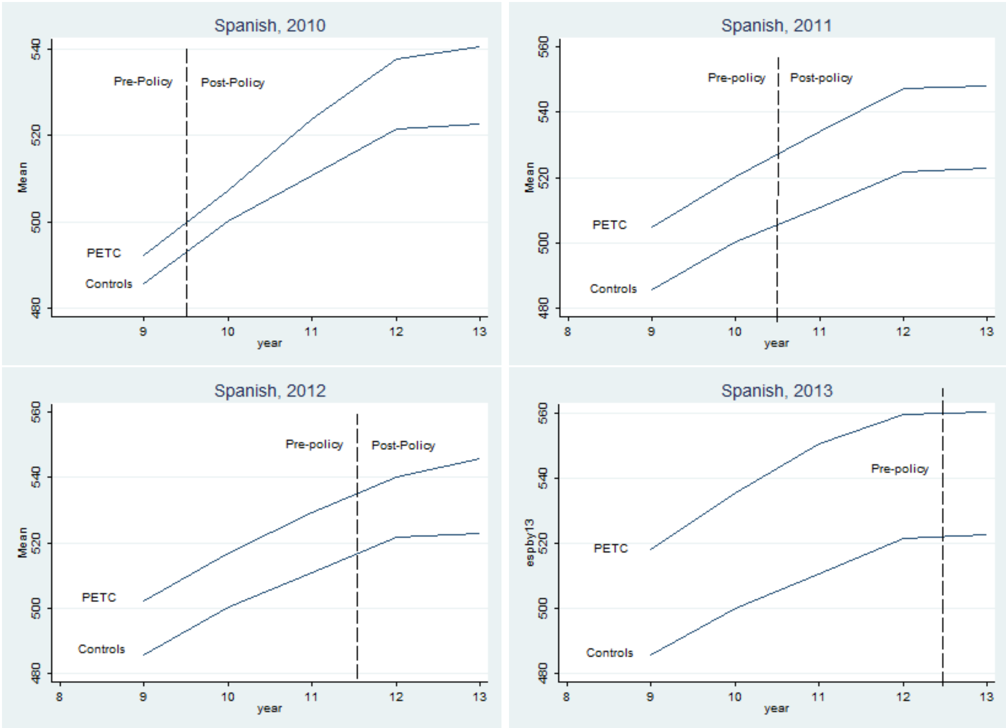


FIGURE B2: Trends of math average scores ENLACE after PSM by treatment status

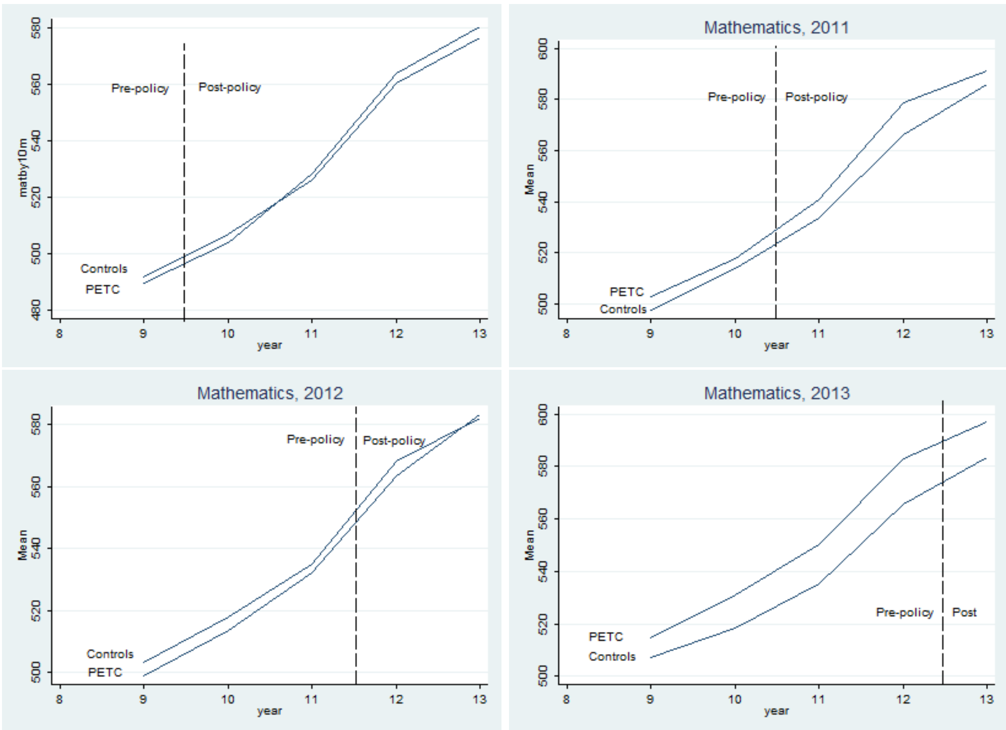


TABLE B2: Differences in Differences: Standardized Mathematics Test Scores by PETC Cohort

	(1) Raw	(2) Controls	(3) Panel	(4) Low Mg	(5) High Mg	(6) No PEC	(7) PEC	(8) Subsample)
One Year of Treatment								
(2009-2010)	-0.006	-0.012	-0.005	-0.019	0.056	0.041	-0.041	-0.002
	(0.022)	(0.023)	(0.022)	(0.022)	(0.058)	(0.040)	(0.027)	(0.022)
N	101958	98647	87631	60245	41441	52525	49161	90251
R^2	0.000	0.290	0.273	0.299	0.196	0.265	0.271	0.281
(2010-2011)	0.036	0.057	0.074*	0.019	0.194*	0.171**	-0.011	0.038
	(0.043)	(0.040)	(0.041)	(0.037)	(0.110)	(0.076)	(0.045)	(0.040)
N	102949	99445	87351	59918	42563	52855	49626	90444
R^2	0.000	0.258	0.243	0.259	0.191	0.249	0.220	0.254
(2011-2012)	-0.032*	-0.013	0.013	-0.016	-0.019	-0.018	-0.026	0.005
	(0.017)	(0.017)	(0.016)	(0.017)	(0.049)	(0.037)	(0.018)	(0.016)
N	99524	95541	87135	58043	40471	50362	48152	89335
R^2	0.008	0.229	0.235	0.219	0.199	0.232	0.176	0.240
(2012-2013)	-0.019	0.096***	0.087***	0.046	0.102	-0.004	0.039	0.092***
	(0.031)	(0.034)	(0.031)	(0.030)	(0.146)	(0.071)	(0.033)	(0.030)
N	97862	94348	87182	57421	39851	49477	47795	88980
R^2	0.002	0.194	0.200	0.165	0.166	0.183	0.129	0.203
Two Years of Treatment								
(2009-2010)	0.040	0.068**	0.080***	0.037	0.180***	0.035	0.065**	0.070***
	(0.026)	(0.027)	(0.027)	(0.027)	(0.063)	(0.046)	(0.032)	(0.027)
N	102195	98548	87346	59557	42022	52496	49083	89987
R^2	0.001	0.272	0.257	0.280	0.200	0.254	0.248	0.267
(2010-2011)	0.107*	0.119*	0.117*	0.029	0.431***	0.194*	0.066	0.109*
	(0.062)	(0.061)	(0.061)	(0.058)	(0.159)	(0.116)	(0.069)	(0.058)
N	98959	95621	87408	58720	39882	50381	48221	89580
R^2	0.007	0.227	0.233	0.222	0.184	0.229	0.179	0.238
(2011-2012)	0.023	0.057***	0.085***	0.022	0.179***	0.032	0.027	0.077***
	(0.018)	(0.019)	(0.018)	(0.019)	(0.053)	(0.040)	(0.021)	(0.019)
N	102196	98191	87137	58632	42538	51959	49211	89863
R^2	0.005	0.184	0.194	0.159	0.144	0.174	0.121	0.202
Three Years of Treatment								
(2009-2010)	0.047	0.089**	0.123***	0.038	0.268***	-0.020	0.112**	0.101***
	(0.037)	(0.039)	(0.038)	(0.038)	(0.098)	(0.069)	(0.044)	(0.038)
N	98195	94725	87403	58360	39341	50023	47678	89124
R^2	0.004	0.229	0.232	0.229	0.180	0.228	0.189	0.239
(2010-2011)	0.072	0.119*	0.142**	0.044	0.300**	0.168	0.037	0.111*
	(0.069)	(0.070)	(0.070)	(0.073)	(0.143)	(0.133)	(0.077)	(0.067)
N	101625	98271	87410	59307	41951	51978	49280	90108
R^2	0.004	0.180	0.191	0.165	0.131	0.171	0.127	0.199
Four Years of Treatment								
(2009-2010)	0.064*	0.133***	0.166***	0.109***	0.206**	0.048	0.122***	0.154***
	(0.038)	(0.040)	(0.039)	(0.039)	(0.095)	(0.068)	(0.047)	(0.039)
N	100870	97375	87405	58946	41411	51620	48737	89652
R^2	0.002	0.186	0.195	0.179	0.130	0.174	0.144	0.204

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Standard errors, clustered on school, in parentheses. Columns 4 to 8 include a full set of ‘X’ controls based on charachteristics of an unbalanced panel of schools. Column 8 excludes the states of Michoacan, Guerrero, Oaxaca and Campeche, since these states have been recently signaled by the Mexican media as not being accountable in their ENLACE results.

TABLE B3: PSM and Differences in Differences: Standardized Mathematics Test Scores by PETC Cohort

	(1) Raw	(2) Controls	(3) Panel	(4) Low Mg	(5) High Mg	(6) No PEC	(7) PEC	(8) Subsample
One Year of Treatment								
(2009-2010)	-0.006	-0.008	-0.010	-0.028	0.086	0.049	-0.028	-0.008
	(0.023)	(0.024)	(0.029)	(0.025)	(0.069)	(0.042)	(0.032)	(0.024)
N	11514	11291	10182	8189	3142	4124	7207	10524
R^2	0.000	0.273	0.256	0.277	0.224	0.273	0.261	0.265
(2010-2011)	0.028	0.049	0.063	0.009	0.192	0.198**	-0.028	0.034
	(0.043)	(0.052)	(0.049)	(0.038)	(0.122)	(0.086)	(0.045)	(0.031)
N	4236	4185	3675	3126	1071	1425	2772	3790
R^2	0.001	0.263	0.256	0.270	0.239	0.270	0.253	0.264
(2011-2012)	-0.029*	-0.003	-0.012	-0.007	-0.020	-0.023	-0.004	-0.009
	(0.016)	(0.018)	(0.014)	(0.019)	(0.062)	(0.034)	(0.019)	(0.016)
N	23505	22972	22117	17209	5907	7183	15933	22351
R^2	0.012	0.204	0.206	0.195	0.196	0.228	0.169	0.207
(2012-2013)	-0.051	0.030	0.038	-0.006	-0.019	-0.055	-0.003	0.051
	(0.031)	(0.035)	(0.032)	(0.033)	(0.167)	(0.076)	(0.030)	(0.033)
N	6033	5708	5429	5142	856	1407	4591	5410
R^2	0.010	0.159	0.162	0.145	0.206	0.217	0.133	0.156
Two Years of Treatment								
(2009-2010)	0.038	0.071**	0.083***	0.030	0.144*	0.021	0.095***	0.070**
	(0.025)	(0.030)	(0.027)	(0.032)	(0.074)	(0.050)	(0.032)	(0.030)
N	11450	11201	10160	8026	3215	4084	7157	10458
R^2	0.001	0.258	0.249	0.266	0.237	0.260	0.252	0.255
(2010-2011)	0.109*	0.108	0.090*	0.006	0.408**	0.202*	0.072	0.088
	(0.064)	(0.066)	(0.052)	(0.057)	(0.168)	(0.122)	(0.083)	(0.069)
N	4038	3987	3675	2990	1009	1315	2684	3737
R^2	0.009	0.238	0.250	0.254	0.230	0.278	0.214	0.252
(2011-2012)	0.037*	0.058***	0.053***	0.029	0.166***	0.031	0.063***	0.052***
	(0.020)	(0.019)	(0.019)	(0.019)	(0.062)	(0.044)	(0.022)	(0.019)
N	23523	22976	22115	17126	5994	7173	15947	22334
R^2	0.007	0.152	0.151	0.129	0.118	0.152	0.106	0.156
Three Years of Treatment								
(2009-2010)	0.063*	0.082*	0.099**	0.039	0.263**	0.006	0.127***	0.074*
	(0.035)	(0.042)	(0.044)	(0.036)	(0.115)	(0.069)	(0.046)	(0.041)
N	11026	10797	10162	7870	2967	3858	6979	10375
R^2	0.006	0.227	0.223	0.236	0.234	0.254	0.204	0.227
(2010-2011)	0.028	0.042	0.057	-0.033	0.108	0.089	-0.008	0.027
	(0.077)	(0.093)	(0.073)	(0.078)	(0.152)	(0.149)	(0.072)	(0.071)
N	4170	4116	3675	3013	1115	1388	2740	3755
R^2	0.002	0.157	0.185	0.131	0.182	0.167	0.148	0.187
Four Years of Treatment								
(2009-2010)	0.064*	0.123***	0.140***	0.101**	0.189*	0.056	0.149***	0.123***
	(0.038)	(0.045)	(0.045)	(0.044)	(0.105)	(0.072)	(0.050)	(0.041)
N	11364	11118	10161	7950	3208	4014	7144	10406
R^2	0.002	0.169	0.177	0.178	0.137	0.172	0.153	0.177

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Standard errors, clustered on school, in parentheses. Columns 4 to 8 include a full set of 'X' controls based on charachteristics of an unbalanced panel of schools. Column 8 excludes the states of Michoacan, Guerrero, Oaxaca and Campeche, since these states have been recently signaled by the Mexican media as not being accountable in their ENLACE results.

TABLE B4: Differences in Differences: Standardized Spanish Test Scores by PETC Cohort

	(1) Raw	(2) Controls	(3) Panel	(4) Low Mg	(5) High Mg	(6) No PEC	(7) PEC	(8) Subsample
One Year of Treatment								
(2009-2010)	0.020	0.001	0.009	-0.006	0.072	0.045	-0.025	0.012
	(0.022)	(0.023)	(0.022)	(0.022)	(0.058)	(0.038)	(0.027)	(0.022)
N	101958	98647	87631	60245	41441	52525	49161	90251
R^2	0.000	0.330	0.312	0.335	0.202	0.288	0.312	0.316
(2010-2011)	0.041	0.060	0.073**	0.020	0.186*	0.145**	0.003	0.037
	(0.041)	(0.037)	(0.037)	(0.036)	(0.097)	(0.067)	(0.043)	(0.036)
N	102949	99445	87351	59918	42563	52855	49626	90444
R^2	0.000	0.302	0.287	0.301	0.197	0.280	0.266	0.293
(2011-2012)	-0.025	-0.004	0.021	-0.020	-0.011	-0.004	-0.021	0.013
	(0.016)	(0.016)	(0.015)	(0.016)	(0.047)	(0.036)	(0.017)	(0.016)
N	99513	95532	87126	58041	40462	50351	48152	89326
R^2	0.011	0.260	0.268	0.245	0.200	0.254	0.206	0.271
(2012-2013)	-0.002	0.110***	0.099***	0.028	0.192	0.048	0.020	0.110***
	(0.031)	(0.035)	(0.032)	(0.030)	(0.139)	(0.070)	(0.033)	(0.030)
N	97852	94340	87173	57420	39842	49467	47795	88972
R^2	0.002	0.219	0.228	0.194	0.158	0.208	0.154	0.231
Two Years of Treatment								
(2009-2010)	0.070***	0.097***	0.107***	0.067***	0.190***	0.068	0.087***	0.100***
	(0.024)	(0.026)	(0.026)	(0.026)	(0.063)	(0.044)	(0.031)	(0.026)
N	102195	98548	87346	59557	42022	52496	49083	89987
R^2	0.001	0.315	0.299	0.321	0.201	0.281	0.293	0.304
(2010-2011)	0.084	0.101*	0.095*	0.010	0.352**	0.151	0.051	0.093*
	(0.057)	(0.053)	(0.053)	(0.052)	(0.137)	(0.101)	(0.060)	(0.051)
N	98948	95612	87399	58718	39873	50370	48221	89571
R^2	0.009	0.259	0.265	0.246	0.186	0.251	0.208	0.269
(2011-2012)	0.027	0.069***	0.096***	0.017	0.168***	0.031	0.034*	0.086***
	(0.017)	(0.018)	(0.018)	(0.018)	(0.052)	(0.038)	(0.020)	(0.018)
N	102197	98192	87137	58633	42538	51960	49211	89864
R^2	0.007	0.218	0.234	0.200	0.145	0.208	0.157	0.239
Three Years of Treatment								
(2009-2010)	0.062*	0.111***	0.140***	0.045	0.280***	-0.002	0.129***	0.122***
	(0.036)	(0.038)	(0.037)	(0.036)	(0.095)	(0.068)	(0.043)	(0.037)
N	98184	94716	87394	58358	39332	50012	47678	89115
R^2	0.006	0.260	0.263	0.253	0.174	0.245	0.219	0.268
(2010-2011)	0.073	0.109*	0.138**	0.026	0.255**	0.154	0.018	0.109*
	(0.065)	(0.063)	(0.066)	(0.068)	(0.129)	(0.116)	(0.073)	(0.061)
N	101626	98272	87410	59308	41951	51979	49280	90109
R^2	0.005	0.216	0.231	0.205	0.131	0.206	0.162	0.236
Four Years of Treatment								
(2009-2010)	0.078**	0.161***	0.188***	0.114***	0.236**	0.041	0.157***	0.181***
	(0.036)	(0.038)	(0.038)	(0.037)	(0.095)	(0.063)	(0.046)	(0.037)
N	100871	97376	87405	58947	41411	51621	48737	89653
R^2	0.003	0.223	0.236	0.220	0.125	0.206	0.183	0.241

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Standard errors, clustered on school, in parentheses. Columns 4 to 8 include a full set of 'X' controls based on charachteristics of an unbalanced panel of schools. Column 8 excludes the states of Michoacan, Guerrero, Oaxaca and Campeche, since these states have been recently signaled by the Mexican media as not being accountable in their ENLACE results.

TABLE B5: PSM and Differences in Differences: Standardized Spanish Test Scores by PETC Cohort

	(1) Raw	(2) Full Sample	(3) Panel	(4) Low Mg	(5) High Mg	(6) No PEC	(7) PEC	(8) Subsample
One Year of Treatment								
(2009-2010)	0.022	0.004	0.000	-0.015	0.097	0.046	-0.009	0.003
	(0.023)	(0.026)	(0.023)	(0.025)	(0.066)	(0.045)	(0.034)	(0.023)
N	11514	11291	10182	8189	3142	4124	7207	10524
R^2	0.000	0.316	0.294	0.313	0.224	0.288	0.305	0.298
(2010-2011)	0.033	0.054**	0.066**	0.011	0.208*	0.163***	-0.010	0.035
	(0.043)	(0.027)	(0.032)	(0.039)	(0.109)	(0.062)	(0.046)	(0.031)
N	4236	4185	3675	3126	1071	1425	2772	3790
R^2	0.001	0.295	0.287	0.291	0.241	0.294	0.285	0.291
(2011-2012)	-0.020	0.004	-0.002	-0.011	0.011	-0.012	0.002	0.001
	(0.017)	(0.018)	(0.016)	(0.016)	(0.051)	(0.037)	(0.018)	(0.018)
N	23504	22971	22116	17209	5906	7182	15933	22350
R^2	0.015	0.227	0.229	0.203	0.209	0.235	0.195	0.231
(2012-2013)	-0.043	0.028	0.032	-0.014	0.074	-0.029	-0.012	0.049*
	(0.035)	(0.035)	(0.033)	(0.028)	(0.165)	(0.074)	(0.034)	(0.027)
N	6033	5708	5429	5142	856	1407	4591	5410
R^2	0.010	0.163	0.172	0.129	0.195	0.212	0.127	0.173
Two Years of Treatment								
(2009-2010)	0.057**	0.086***	0.095***	0.053*	0.134*	0.031	0.108***	0.085***
	(0.029)	(0.027)	(0.028)	(0.028)	(0.071)	(0.045)	(0.033)	(0.026)
N	11450	11201	10160	8026	3215	4084	7157	10458
R^2	0.002	0.299	0.288	0.300	0.235	0.273	0.298	0.290
(2010-2011)	0.079	0.082	0.062	-0.015	0.342**	0.116	0.058	0.065
	(0.053)	(0.061)	(0.056)	(0.057)	(0.143)	(0.107)	(0.054)	(0.055)
N	4038	3987	3675	2990	1009	1315	2684	3737
R^2	0.009	0.258	0.270	0.256	0.233	0.275	0.239	0.266
(2011-2012)	0.021	0.051***	0.047**	0.019	0.149***	-0.003	0.063***	0.046***
	(0.019)	(0.017)	(0.019)	(0.016)	(0.049)	(0.033)	(0.020)	(0.017)
N	23523	22976	22115	17126	5994	7173	15947	22334
R^2	0.007	0.175	0.176	0.151	0.123	0.176	0.138	0.181
Three Years of Treatment								
(2009-2010)	0.063*	0.091**	0.100**	0.038	0.267**	0.007	0.130***	0.081**
	(0.036)	(0.038)	(0.040)	(0.038)	(0.124)	(0.070)	(0.048)	(0.040)
N	11025	10796	10161	7870	2966	3857	6979	10374
R^2	0.007	0.254	0.246	0.254	0.223	0.253	0.232	0.249
(2010-2011)	0.024	0.034	0.059	-0.035	0.118	0.062	-0.011	0.032
	(0.074)	(0.065)	(0.076)	(0.088)	(0.145)	(0.136)	(0.081)	(0.070)
N	4170	4116	3675	3013	1115	1388	2740	3755
R^2	0.001	0.184	0.218	0.151	0.176	0.193	0.182	0.215
Four Years of Treatment								
(2009-2010)	0.064	0.132***	0.143***	0.099**	0.218**	0.028	0.173***	0.133***
	(0.042)	(0.035)	(0.044)	(0.042)	(0.104)	(0.068)	(0.053)	(0.037)
N	11364	11118	10161	7950	3208	4014	7144	10406
R^2	0.002	0.203	0.211	0.210	0.127	0.190	0.191	0.209

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Standard errors, clustered on school, in parentheses. Columns 4 to 8 include a full set of 'X' controls based on charachteristics of an unbalanced panel of schools. Column 8 excludes the states of Michoacan, Guerrero, Oaxaca and Campeche, since these states have been recently signaled by the Mexican media as not being accountable in their ENLACE results.

FIGURE B3: Trends of Spanish average scores ENLACE after PSM by treatment status

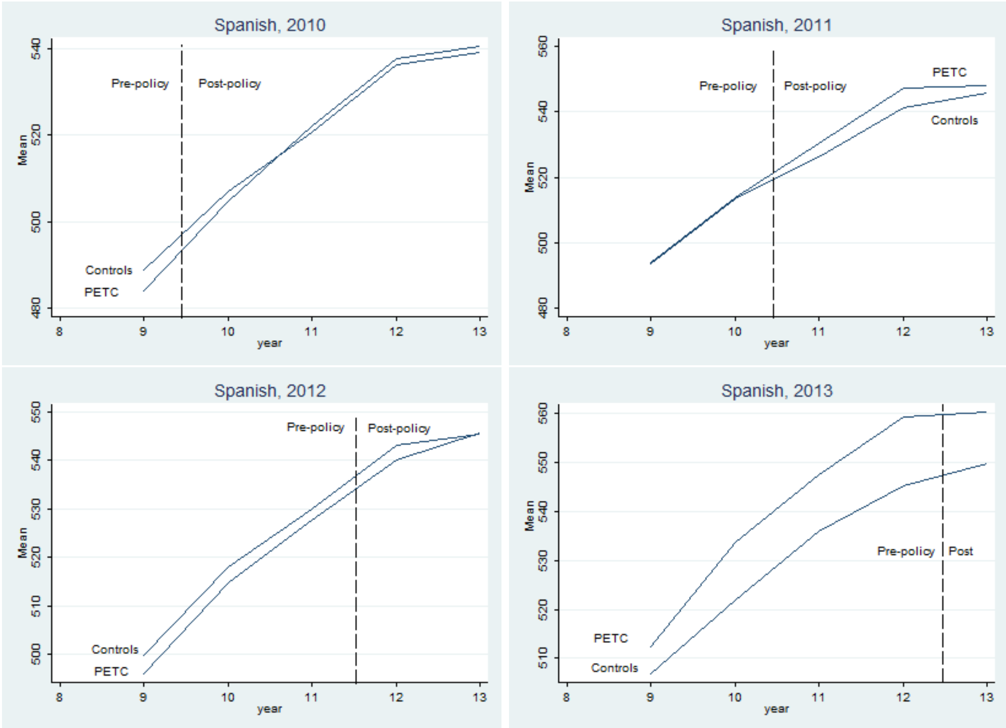


TABLE B6: Placebo regressions: DiD and PSM-DiD for dropout rates

	DiD			*PSM and DiD*		
	PETC 2011	PETC 2012	PETC 2013	PETC 2011	PETC 2012	PETC 2013
One Year Before Policy	0.219	-0.200	-0.193	0.139	-0.333	-0.146
	(0.591)	(0.195)	(0.434)	(0.645)	(0.228)	(0.482)
N	65327	63857	60245	2594	13710	2881
Two Years Before Policy		0.038	0.954**		-0.040	0.630
		(0.181)	(0.398)		(0.178)	(0.425)
N		65338	63846		14207	2933
Three Years Before Policy			0.044			-0.059
			(0.339)			(0.311)
N			65324			3067

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
DiD regressions show standard errors, clustered on school, in parenthesis. PSM and DiD regressions show robust standard errors from 100 replications, 100% of replacement and clustered on school, in parentheses. Regressions include a full set of controls, including school's teachers' and principals' characteristics as well as controls for the number of years in PEC, marginality of the school area and dummies for six mexican regions.

TABLE B7: Probability of schools being treated, 2010

Variables	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
PEC	0.14934	0.04292	3.48	0.0010	0.06523	0.23346
# students	0.00026	0.00289	0.09	0.9280	-0.00540	0.00593
# students squared	-0.00001	0.00001	-1.33	0.1830	-0.00002	0.00000
# principals	0.21756	0.14212	1.53	0.1260	-0.06100	0.49611
# teachers	-0.44947	0.26640	-1.69	0.0920	-0.97162	0.07267
# teachers square	0.00882	0.00190	4.65	0.0000	0.00511	0.01254
# administrative workers	0.07149	0.03566	2.00	0.0450	0.00160	0.14138
# principals vocational	-0.29247	0.12395	-2.36	0.0180	-0.53542	-0.04953
# principals bachelors	-0.26640	0.12340	-2.16	0.0310	-0.50827	-0.02454
# principals postgraduate	-0.21370	0.13031	-1.64	0.1010	-0.46911	0.04171
# teachers vocational	0.44983	0.26380	1.71	0.0880	-0.06720	0.96687
# teachers bachelors	0.45577	0.26367	1.73	0.0840	-0.06101	0.97255
# teachers postgraduate	0.29498	0.26619	1.11	0.2680	-0.22674	0.81670
# secretary	-0.01411	0.08360	-0.17	0.8660	-0.17797	0.14975
# deputy administrative	-0.32496	0.21062	-1.54	0.1230	-0.73777	0.08785
# cleaning personel	-0.09972	0.05794	-1.72	0.0850	-0.21329	0.01385
# janitors	-0.04656	0.08170	-0.57	0.5690	-0.20670	0.11357
hours instruction sports	0.00481	0.00394	1.22	0.2220	-0.00290	0.01253
hours instruction arts	-0.00831	0.01026	-0.81	0.4180	-0.02842	0.01180
hours instruction IT	0.01145	0.01000	1.15	0.2520	-0.00814	0.03105
hours instruction English	0.02799	0.00640	4.37	0.0000	0.01545	0.04054
# teachers “carrera magisterial”	-0.00066	0.01128	-0.06	0.9530	-0.02276	0.02144
# classrooms	-0.00301	0.01278	-0.24	0.8140	-0.02805	0.02203
# classrooms per grade	-0.10560	0.02834	-3.73	0.0000	-0.16114	-0.05006
# classrooms per grade (adapted)	-0.00266	0.02235	-0.12	0.9050	-0.04647	0.04115
average spending in books	0.00003	0.00001	1.79	0.0730	0.00000	0.00006
average spending in uniforms	-0.00002	0.00005	-0.38	0.7030	-0.00013	0.00009
average spending in fees	-0.00028	0.00015	-1.96	0.0500	-0.00057	0.00000
students tested	0.00456	0.00437	1.04	0.2960	-0.00400	0.01313
students tested squared	-0.00001	0.00001	-0.86	0.3920	-0.00004	0.00002
# of untrusted tests	0.01629	0.00537	3.03	0.0020	0.00576	0.02682
# of untrusted tests squared	-0.00030	0.00014	-2.14	0.0320	-0.00058	-0.00003
Marinality index 2	-0.11280	0.05942	-1.90	0.0580	-0.22927	0.00367
Marginality Index 3	-0.28312	0.06778	-4.18	0.0000	-0.41596	-0.15027
Marginality index 4	-0.49965	0.06618	-7.55	0.0000	-0.62936	-0.36994
Marginality index 5	-0.57769	0.11987	-4.82	0.0000	-0.81264	-0.34274
Region 2	-0.13618	0.06859	-1.99	0.0470	-0.27062	-0.00174
Region 3	-0.75196	0.07784	-9.66	0.0000	-0.90452	-0.59941
Region 4	-0.80797	0.10101	-8.00	0.0000	-1.00594	-0.60999
Region 5	-0.41425	0.06378	-6.49	0.0000	-0.53926	-0.28924
Region 6	-0.29368	0.06437	-4.56	0.0000	-0.41983	-0.16753
Constant	-1.48603	0.11583	-12.83	0.0000	-1.71305	-1.25901

*Propensity score matching using 34165 observations in 2010. Prob > chi2 is equal to 0.0000 and Pseudo R2 = 0.1028. Marginality index 1 and Region 1 are ommited

TABLE B8: Probability of schools being treated, 2011

Variables	Coef.	Std. Err.	z	P>Z	[95% Conf.	Interval]
PEC	0.20229	0.06504	3.11	0.0020	0.07480	0.32977
# students	0.00441	0.00369	1.20	0.2320	-0.00282	0.01164
# students squared	0.00000	0.00001	-0.12	0.9030	-0.00002	0.00001
# principals	-3.45948	92.69351	-0.04	0.9700	-185.13540	178.21650
# teachers	-0.12488	0.30540	-0.41	0.6830	-0.72346	0.47370
# teachers square	-0.00523	0.00347	-1.51	0.1320	-0.01202	0.00157
# administrative workers	0.06718	0.05518	1.22	0.2230	-0.04097	0.17533
# principals vocational	3.18948	92.69344	0.03	0.9730	-178.48630	184.86530
# principals bachelors	3.35250	92.69344	0.04	0.9710	-178.32330	185.02830
# principals postgraduate	3.38726	92.69344	0.04	0.9710	-178.28860	185.06310
# teachers vocational	0.16549	0.29929	0.55	0.5800	-0.42111	0.75210
# teachers bachelors	0.18364	0.29900	0.61	0.5390	-0.40238	0.76966
# teachers postgraduate	0.05191	0.30389	0.17	0.8640	-0.54370	0.64752
# secretary	0.05513	0.11532	0.48	0.6330	-0.17088	0.28115
# deputy administrative	-0.24102	0.27492	-0.88	0.3810	-0.77985	0.29781
# cleaning personel	-0.17867	0.08886	-2.01	0.0440	-0.35283	-0.00451
# janitors	-0.18449	0.12225	-1.51	0.1310	-0.42409	0.05512
hours instruction sports	0.00622	0.00568	1.10	0.2730	-0.00490	0.01734
hours instruction arts	0.02232	0.01021	2.19	0.0290	0.00232	0.04232
hours instruction IT	0.00927	0.01109	0.84	0.4030	-0.01246	0.03100
hours instruction English	0.00575	0.00806	0.71	0.4750	-0.01004	0.02154
# teachers “carrera magisterial”	0.02076	0.01722	1.21	0.2280	-0.01298	0.05450
# classrooms	0.01539	0.01900	0.81	0.4180	-0.02185	0.05263
# classrooms per grade	-0.01485	0.05216	-0.28	0.7760	-0.11708	0.08738
# classrooms per grade (adapted)	-0.04384	0.04310	-1.02	0.3090	-0.12832	0.04064
average spending in books	-0.00006	0.00013	-0.47	0.6420	-0.00031	0.00019
average spending in uniforms	0.00002	0.00006	0.26	0.7980	-0.00010	0.00013
average spending in fees	0.00001	0.00005	0.28	0.7820	-0.00008	0.00011
students tested	-0.00458	0.00566	-0.81	0.4180	-0.01566	0.00651
students tested squared	-0.00001	0.00002	-0.34	0.7320	-0.00004	0.00003
# of untrusted tests	0.01519	0.00868	1.75	0.0800	-0.00181	0.03220
# of untrusted tests squared	-0.00031	0.00024	-1.29	0.1970	-0.00077	0.00016
Marinality index 2	-0.34220	0.09613	-3.56	0.0000	-0.53060	-0.15379
Marginality Index 3	-0.26440	0.10217	-2.59	0.0100	-0.46464	-0.06416
Marginality index 4	-0.45650	0.10213	-4.47	0.0000	-0.65668	-0.25633
Marginality index 5	-0.36308	0.14876	-2.44	0.0150	-0.65465	-0.07150
Region 2	0.25336	0.10642	2.38	0.0170	0.04477	0.46195
Region 3	-0.16267	0.10796	-1.51	0.1320	-0.37427	0.04893
Region 4	-0.31072	0.14050	-2.21	0.0270	-0.58610	-0.03534
Region 5	0.09216	0.09320	0.99	0.3230	-0.09051	0.27484
Region 6	-0.03362	0.10887	-0.31	0.7570	-0.24701	0.17976
Constant	-2.54411	0.18488	-13.76	0.0000	-2.90645	-2.18176

*Propensity score matching using 33084 observations in 2010. Prob > chi2 is equal to 0.0000 and Pseudo R2 = 0.0889, marginality index 1 and Region 1 are ommited

TABLE B9: Probability of schools being treated, 2012

Variables	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
PEC	0.24902	0.03123	7.97	0.0000	0.18782	0.31023
# students	0.00002	0.00191	0.01	0.9930	-0.00373	0.00376
# students squared	0.00000	0.00000	0.06	0.9480	-0.00001	0.00001
# principals	-0.07586	0.13495	-0.56	0.5740	-0.34036	0.18865
# teachers	0.07504	0.10515	0.71	0.4750	-0.13105	0.28112
# teachers square	-0.01333	0.00164	-8.14	0.0000	-0.01655	-0.01012
# administrative workers	0.14577	0.02423	6.02	0.0000	0.09828	0.19327
# principals vocational	0.20232	0.12252	1.65	0.0990	-0.03782	0.44245
# principals bachelors	0.18437	0.12189	1.51	0.1300	-0.05453	0.42327
# principals postgraduate	0.17432	0.12461	1.40	0.1620	-0.06991	0.41855
# teachers vocational	0.09653	0.10120	0.95	0.3400	-0.10181	0.29487
# teachers bachelors	0.12777	0.10089	1.27	0.2050	-0.06997	0.32551
# teachers postgraduate	0.10947	0.10247	1.07	0.2850	-0.09136	0.31030
# secretary	-0.16655	0.05944	-2.80	0.0050	-0.28304	-0.05005
# deputy administrative	-0.11199	0.07160	-1.56	0.1180	-0.25232	0.02835
# cleaning personel	-0.19033	0.04035	-4.72	0.0000	-0.26941	-0.11125
# janitors	-0.07756	0.05315	-1.46	0.1450	-0.18174	0.02662
hours instruction sports	-0.00385	0.00275	-1.40	0.1620	-0.00923	0.00154
hours instruction arts	0.01417	0.00598	2.37	0.0180	0.00246	0.02588
hours instruction IT	0.01872	0.00604	3.10	0.0020	0.00688	0.03056
hours instruction English	0.01718	0.00360	4.77	0.0000	0.01011	0.02424
# teachers “carrera magisterial”	0.00511	0.00812	0.63	0.5290	-0.01080	0.02101
# classrooms	0.03664	0.00926	3.95	0.0000	0.01848	0.05480
# classrooms per grade	0.00820	0.02521	0.33	0.7450	-0.04120	0.05761
# classrooms per grade (adapted)	0.00989	0.01484	0.67	0.5050	-0.01920	0.03898
average spending in books	0.00000	0.00003	0.10	0.9170	-0.00006	0.00006
average spending in uniforms	-0.00018	0.00006	-3.26	0.0010	-0.00030	-0.00007
average spending in fees	0.00002	0.00003	0.49	0.6270	-0.00005	0.00008
students tested	0.00136	0.00287	0.47	0.6350	-0.00427	0.00699
students tested squared	-0.00001	0.00001	-1.01	0.3110	-0.00002	0.00001
# of untrusted tests	-0.00430	0.00291	-1.48	0.1390	-0.00999	0.00140
# of untrusted tests squared	0.00000	0.00005	0.03	0.9800	-0.00009	0.00009
Marinality index 2	0.02195	0.04391	0.50	0.6170	-0.06411	0.10800
Marginality Index 3	-0.16590	0.05305	-3.13	0.0020	-0.26989	-0.06192
Marginality index 4	-0.20521	0.04881	-4.20	0.0000	-0.30088	-0.10954
Marginality index 5	-0.08946	0.09226	-0.97	0.3320	-0.27028	0.09136
Region 2	0.33088	0.07021	4.71	0.0000	0.19326	0.46850
Region 3	-0.18247	0.04734	-3.85	0.0000	-0.27527	-0.08968
Region 4	-0.68344	0.08333	-8.20	0.0000	-0.84676	-0.52013
Region 5	0.05558	0.04470	1.24	0.2140	-0.03203	0.14319
Region 6	-0.09418	0.05220	-1.80	0.0710	-0.19649	0.00813
Constant	-2.68645	0.10290	-26.11	0.0000	-2.88813	-2.48478

*Propensity score matching using 30710 observations in 2010. Prob > chi2 is equal to 0.0000 and Pseudo R2 = 0.1277, marginality index 1 and Region 1 are ommited

TABLE B10: Probability of schools being treated, 2013

Variables	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
PEC	0.15963	0.06728	2.37	0.0180	0.02776	0.29150
# students	0.00566	0.00361	1.57	0.1170	-0.00142	0.01274
# students squared	-0.00001	0.00001	-1.23	0.2180	-0.00002	0.00001
# principals	-0.00836	0.28201	-0.03	0.9760	-0.56108	0.54436
# teachers	0.33171	0.13147	2.52	0.0120	0.07402	0.58939
# teachers square	-0.01369	0.00366	-3.74	0.0000	-0.02087	-0.00651
# administrative workers	0.04298	0.03091	1.39	0.1640	-0.01760	0.10356
# principals vocational	-0.02333	0.22855	-0.10	0.9190	-0.47128	0.42461
# principals bachelors	-0.00838	0.22659	-0.04	0.9710	-0.45249	0.43573
# principals postgraduate	-0.08975	0.23208	-0.39	0.6990	-0.54462	0.36512
# teachers vocational	-0.05017	0.11211	-0.45	0.6540	-0.26990	0.16956
# teachers bachelors	-0.03252	0.11123	-0.29	0.7700	-0.25053	0.18549
# teachers postgraduate	-0.05551	0.11596	-0.48	0.6320	-0.28278	0.17176
# secretary	0.11890	0.08029	1.48	0.1390	-0.03846	0.27626
# deputy administrative	-0.01220	0.13533	-0.09	0.9280	-0.27744	0.25305
# cleaning personel	0.10015	0.06259	1.60	0.1100	-0.02252	0.22282
# janitors	0.18526	0.09835	1.88	0.0600	-0.00750	0.37802
hours instruction sports	-0.00360	0.00555	-0.65	0.5170	-0.01448	0.00728
hours instruction arts	0.02086	0.01088	1.92	0.0550	-0.00046	0.04218
hours instruction IT	0.00764	0.01089	0.70	0.4830	-0.01371	0.02898
hours instruction English	0.03421	0.00599	5.71	0.0000	0.02246	0.04595
# teachers “carrera magisterial”	-0.04870	0.01760	-2.77	0.0060	-0.08320	-0.01421
# classrooms	0.03586	0.01787	2.01	0.0450	0.00084	0.07089
# classrooms per grade	0.00003	0.04655	0.00	1.0000	-0.09122	0.09127
# classrooms per grade (adapted)	-0.00186	0.03090	-0.06	0.9520	-0.06242	0.05871
average spending in books	-0.00027	0.00016	-1.67	0.0950	-0.00058	0.00005
average spending in uniforms	0.00005	0.00005	0.97	0.3310	-0.00005	0.00014
average spending in fees	0.00003	0.00004	0.79	0.4290	-0.00005	0.00011
students tested	-0.01181	0.00488	-2.42	0.0150	-0.02137	-0.00225
students tested squared	0.00002	0.00002	1.16	0.2470	-0.00001	0.00005
# of untrusted tests	0.00707	0.00735	0.96	0.3360	-0.00733	0.02148
# of untrusted tests squared	0.00006	0.00014	0.45	0.6540	-0.00021	0.00034
Marinality index 2	0.24862	0.09156	2.72	0.0070	0.06916	0.42809
Marginality Index 3	0.08242	0.12041	0.68	0.4940	-0.15357	0.31841
Marginality index 4	0.11755	0.10720	1.10	0.2730	-0.09255	0.32765
Marginality index 5	-0.10298	0.34415	-0.30	0.7650	-0.77749	0.57154
Region 2	0.00000	(omitted)				
Region 3	-1.25276	0.29562	-4.24	0.0000	-1.83216	-0.67336
Region 4	0.13380	0.09902	1.35	0.1770	-0.06028	0.32788
Region 5	-0.67754	0.13684	-4.95	0.0000	-0.94575	-0.40934
Region 6	0.14215	0.09493	1.50	0.1340	-0.04391	0.32822
Constant	-3.73173	0.27565	-13.54	0.0000	-4.27199	-3.19147

*Propensity score matching using 29172 observations in 2010. Prob > chi2 is equal to 0.0000 and Pseudo R2 = 0.2111. Marginality index 1 and Region 1 are ommited

TABLE B11: Balance test for treated and matched controls in 2010

	Mean			t-test	
Variable	Treated	Control	% of bias	t	p-value
PEC	0.373	0.373	0.00	0.01	0.9950
# students	143.060	151.160	-7.20	-1.28	0.2010
# students squared	29427.000	32528.000	-6.00	-1.23	0.2200
# principals	0.968	0.980	-4.30	-0.69	0.4920
# teachers	5.412	5.711	-7.80	-1.23	0.2200
# teachers square	42.601	46.992	-8.20	-1.26	0.2070
# administrative workers	0.927	1.038	-9.10	-1.23	0.2180
# principals vocational	0.371	0.374	-0.50	-0.07	0.9420
# principals bachelors	0.416	0.413	0.60	0.10	0.9220
# principals postgraduate	0.150	0.158	-2.10	-0.31	0.7590
# teachers vocational	2.234	2.327	-3.80	-0.57	0.5670
# teachers bachelors	2.991	3.181	-6.90	-1.07	0.2870
# teachers postgraduate	0.159	0.168	-1.60	-0.30	0.7670
# secretary	0.075	0.085	-3.40	-0.48	0.6280
# deputy administrative	0.006	0.007	-0.80	-0.15	0.8780
# cleaning personel	0.472	0.513	-6.20	-0.93	0.3540
# janitors	0.097	0.107	-3.40	-0.48	0.6310
hours instruction sports	4.176	4.692	-8.50	-1.23	0.2180
hours instruction arts	0.436	0.491	-2.50	-0.37	0.7100
hours instruction IT	0.418	0.506	-4.20	-0.55	0.5840
hours instruction English	1.193	1.489	-10.10	-1.15	0.2480
# teachers "carrera magisterial"	2.530	2.740	-8.10	-1.22	0.2230
# classrooms	6.693	6.958	-7.10	-1.11	0.2670
# classrooms per grade	5.682	5.949	-8.00	-1.28	0.2000
# classrooms per grade (adapted)	0.240	0.230	1.10	0.17	0.8670
average spending in books	340.380	339.660	0.00	0.01	0.9960
average spending in uniforms	337.920	343.230	-0.50	-0.27	0.7860
average spending in fees	154.700	157.850	-0.60	-0.33	0.7440
students tested	88.313	93.330	-7.10	-1.25	0.2110
students tested squared	11387.000	12578.000	-5.80	-1.19	0.2350
# of untrusted tests	5.384	5.612	-2.70	-0.42	0.6740
# of untrusted tests squared	96.084	100.530	-1.20	-0.23	0.8210
Marinality index 2	0.238	0.227	2.90	0.42	0.6760
Marginality Index 3	0.161	0.150	2.90	0.44	0.6580
Marginality index 4	0.253	0.243	2.20	0.36	0.7220
Marginality index 5	0.030	0.032	-1.10	-0.19	0.8510
Region 2	0.114	0.122	-2.80	-0.41	0.6850
Region 3	0.054	0.071	-5.30	-1.12	0.2620
Region 4	0.028	0.037	-3.70	-0.79	0.4280
Region 5	0.133	0.137	-1.10	-0.16	0.8710
Region 6	0.131	0.129	0.50	0.08	0.9380

TABLE B12: Balance test for treated and matched controls in 2011

	Mean			t-test	
Variable	Treated	Control	% of bias	t	p-value
PEC	0.469	0.490	-4.40	-0.35	0.7250
# students	164.010	165.280	-1.10	-0.10	0.9170
# students squared	37608.000	38260.000	-1.20	-0.12	0.9060
# principals	0.959	0.965	-2.30	-0.21	0.8370
# teachers	5.786	5.785	0.00	0.00	0.9970
# teachers square	44.269	44.128	0.30	0.03	0.9780
# administrative workers	1.083	1.066	1.10	0.09	0.9300
# principals vocational	0.248	0.239	1.90	0.17	0.8630
# principals bachelors	0.503	0.499	0.80	0.07	0.9450
# principals postgraduate	0.207	0.226	-5.10	-0.40	0.6920
# teachers vocational	1.855	1.801	2.50	0.22	0.8290
# teachers bachelors	3.690	3.715	-0.90	-0.08	0.9340
# teachers postgraduate	0.234	0.261	-4.00	-0.37	0.7090
# secretary	0.097	0.106	-3.00	-0.20	0.8410
# deputy administrative	0.014	0.014	0.00	0.00	1.0000
# cleaning personel	0.497	0.515	-2.80	-0.24	0.8100
# janitors	0.103	0.096	2.20	0.18	0.8600
hours instruction sports	5.641	5.688	-0.70	-0.06	0.9530
hours instruction arts	1.172	1.106	2.00	0.15	0.8830
hours instruction IT	0.800	0.850	-1.80	-0.12	0.9020
hours instruction English	1.510	1.859	-9.80	-0.66	0.5090
# teachers "carrera magisterial"	2.614	2.590	0.90	0.08	0.9390
# classrooms	7.262	7.310	-1.40	-0.12	0.9030
# classrooms per grade	6.152	6.192	-1.20	-0.12	0.9080
# classrooms per grade (adapted)	0.172	0.193	-3.00	-0.29	0.7750
average spending in books	262.690	257.080	0.80	0.16	0.8740
average spending in uniforms	369.620	351.980	2.20	0.48	0.6310
average spending in fees	201.720	175.090	5.20	0.93	0.3510
students tested	100.410	101.420	-1.40	-0.13	0.8940
students tested squared	14223.000	14552.000	-1.50	-0.15	0.8780
# of untrusted tests	5.635	5.296	4.10	0.36	0.7190
# of untrusted tests squared	101.230	86.329	4.50	0.50	0.6190
Marinality index 2	0.138	0.159	-5.90	-0.51	0.6100
Marginality Index 3	0.145	0.138	1.90	0.17	0.8670
Marginality index 4	0.214	0.192	4.90	0.47	0.6420
Marginality index 5	0.062	0.070	-3.10	-0.26	0.7950
Region 2	0.159	0.156	0.80	0.06	0.9490
Region 3	0.103	0.091	3.40	0.36	0.7220
Region 4	0.048	0.050	-0.50	-0.05	0.9570
Region 5	0.255	0.273	-4.60	-0.35	0.7300
Region 6	0.124	0.130	-1.90	-0.16	0.8750

TABLE B13: Balance test for treated and matched controls in 2012

	Mean			t-test	
Variable	Treated	Control	% of bias	t	p-value
PEC	0.564	0.573	-1.90	-0.43	0.6640
# students	181.710	184.180	-2.10	-0.56	0.5770
# students squared	43284.000	44418.000	-2.00	-0.53	0.5980
# principals	1.009	1.010	-0.70	-0.18	0.8540
# teachers	6.595	6.677	-2.30	-0.60	0.5460
# teachers square	53.118	54.471	-2.60	-0.66	0.5100
# administrative workers	1.591	1.490	5.50	1.07	0.2850
# principals vocational	0.345	0.346	-0.20	-0.04	0.9710
# principals bachelors	0.441	0.439	0.30	0.07	0.9460
# principals postgraduate	0.209	0.212	-0.80	-0.16	0.8710
# teachers vocational	1.841	1.864	-1.10	-0.26	0.7930
# teachers bachelors	4.286	4.335	-1.70	-0.42	0.6780
# teachers postgraduate	0.440	0.451	-1.30	-0.29	0.7720
# secretary	0.072	0.072	0.00	0.01	0.9940
# deputy administrative	0.027	0.030	-1.40	-0.28	0.7810
# cleaning personel	0.623	0.619	0.50	0.11	0.9140
# janitors	0.149	0.146	0.90	0.18	0.8600
hours instruction sports	5.577	5.588	-0.20	-0.04	0.9690
hours instruction arts	0.897	0.964	-2.70	-0.51	0.6130
hours instruction IT	0.724	0.908	-7.70	-1.32	0.1870
hours instruction English	2.145	2.207	-1.40	-0.27	0.7880
# teachers “carrera magisterial”	2.849	2.911	-2.40	-0.55	0.5800
# classrooms	8.053	8.097	-1.20	-0.30	0.7630
# classrooms per grade	6.867	6.984	-3.60	-0.92	0.3580
# classrooms per grade (adapted)	0.268	0.288	-2.10	-0.43	0.6670
average spending in books	262.840	274.370	-1.90	-0.47	0.6360
average spending in uniforms	339.930	340.330	-0.10	-0.03	0.9750
average spending in fees	197.730	207.800	-2.40	-0.41	0.6790
students tested	114.610	116.260	-2.20	-0.58	0.5600
students tested squared	17329.000	17792.000	-2.00	-0.53	0.5960
# of untrusted tests	5.549	5.837	-2.70	-0.65	0.5170
# of untrusted tests squared	131.160	141.740	-1.50	-0.45	0.6500
Marinality index 2	0.228	0.222	1.60	0.34	0.7310
Marginality Index 3	0.117	0.118	-0.50	-0.13	0.8980
Marginality index 4	0.233	0.238	-1.10	-0.28	0.7780
Marginality index 5	0.030	0.029	0.70	0.18	0.8580
Region 2	0.057	0.068	-4.90	-1.06	0.2890
Region 3	0.134	0.127	1.70	0.44	0.6600
Region 4	0.049	0.032	6.40	1.97	0.0490
Region 5	0.245	0.262	-4.60	-0.93	0.3520
Region 6	0.113	0.118	-1.70	-0.39	0.6930

TABLE B14: Balance test for treated and matched controls in 2013

	Mean			t-test	
Variable	Treated	Control	% of bias	t	p-value
PEC	0.519	0.556	-7.70	-0.66	0.5120
# students	199.080	200.170	-0.90	-0.09	0.9250
# students squared	49645.000	50670.000	-1.70	-0.17	0.8640
# principals	1.013	1.010	1.30	0.15	0.8830
# teachers	7.577	7.511	1.90	0.19	0.8470
# teachers square	66.218	65.635	1.10	0.10	0.9210
# administrative workers	1.968	2.092	-7.80	-0.51	0.6070
# principals vocational	0.359	0.351	1.80	0.15	0.8790
# principals bachelors	0.436	0.479	-8.50	-0.74	0.4600
# principals postgraduate	0.192	0.164	7.30	0.65	0.5180
# teachers vocational	2.224	2.146	3.60	0.30	0.7650
# teachers bachelors	4.801	4.792	0.30	0.03	0.9770
# teachers postgraduate	0.494	0.526	-3.70	-0.30	0.7630
# secretary	0.231	0.230	0.20	0.01	0.9920
# deputy administrative	0.038	0.035	2.00	0.17	0.8640
# cleaning personel	0.936	0.926	1.40	0.11	0.9120
# janitors	0.179	0.213	-9.00	-0.64	0.5230
hours instruction sports	6.199	6.535	-5.20	-0.43	0.6670
hours instruction arts	1.583	1.425	4.70	0.33	0.7440
hours instruction IT	1.340	1.349	-0.30	-0.02	0.9850
hours instruction English	4.455	4.412	0.80	0.05	0.9580
# teachers “carrera magisterial”	2.590	2.573	0.70	0.07	0.9470
# classrooms	9.128	9.247	-3.40	-0.31	0.7530
# classrooms per grade	7.763	7.749	0.40	0.04	0.9670
# classrooms per grade (adapted)	0.276	0.308	-3.20	-0.23	0.8150
average spending in books	239.540	240.500	-0.20	-0.04	0.9660
average spending in uniforms	431.900	464.050	-4.20	-0.30	0.7660
average spending in fees	232.440	201.640	5.50	0.76	0.4470
students tested	122.940	124.050	-1.50	-0.15	0.8830
students tested squared	19423.000	19856.000	-1.80	-0.17	0.8620
# of untrusted tests	5.000	5.034	-0.40	-0.04	0.9720
# of untrusted tests squared	102.060	91.910	3.40	0.26	0.7940
Marinality index 2	0.372	0.388	-3.70	-0.30	0.7630
Marginality Index 3	0.115	0.110	1.70	0.16	0.8720
Marginality index 4	0.224	0.206	4.10	0.40	0.6910
Marginality index 5	0.006	0.004	1.30	0.23	0.8180
Region 2	0.000	0.000	.	.	.
Region 3	0.006	0.006	0.00	0.00	1.0000
Region 4	0.205	0.210	-1.40	-0.11	0.9110
Region 5	0.051	0.060	-2.90	-0.32	0.7490
Region 6	0.244	0.265	-5.60	-0.44	0.6600

Annex tables and figures: Chapter 3

TABLE C1: Average Total Intra- and Inter-course Dropout Rates

	Dropout rates on 1st to 3rd grades		
School Year	*Total*	*Intra-course*	*Inter-course*
2006-2007	3.14	2.06	1.00
2007-2008	2.59	1.81	0.77
2008-2009	2.53	1.90	0.57
2009-2010	1.82	1.36	0.39
2010-2011	1.76	1.30	0.36
2011-2012	1.03	0.56	0.51
2012-2013	0.87	0.33	0.47
2013-2014	0.06	-0.38	0.39

FIGURE C1: Repetition trends in Mexico City and the rest of Mexico in the pre- and post-policy period.

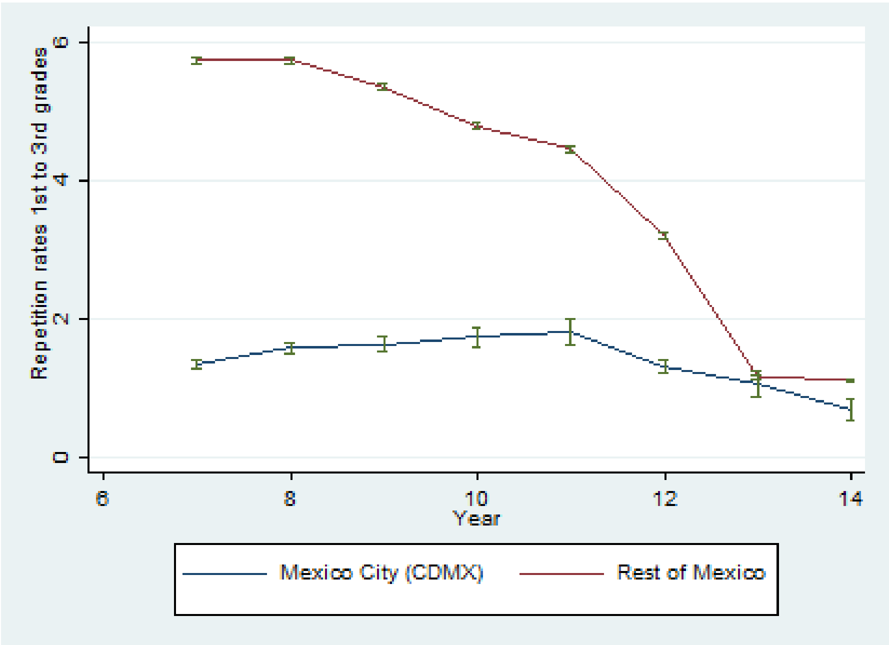


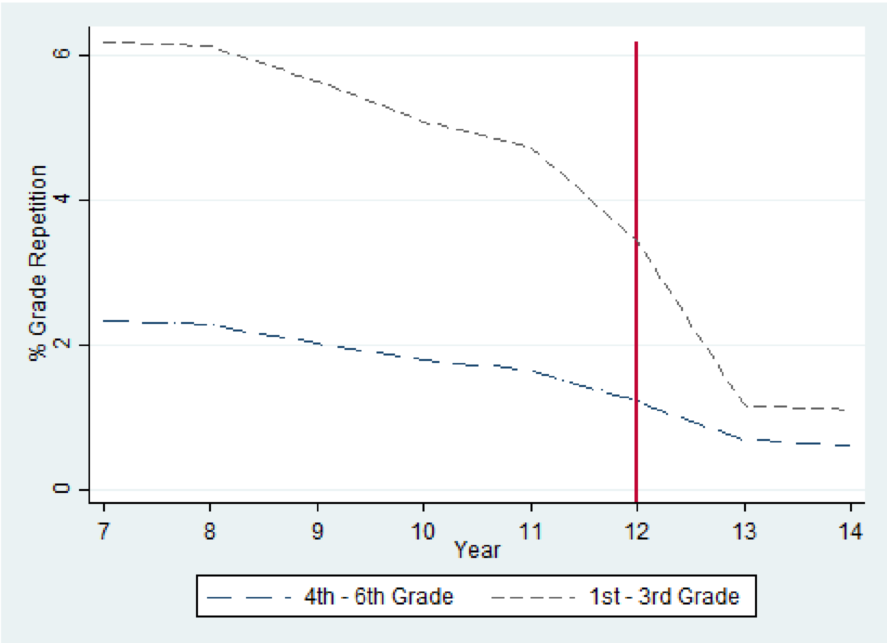
TABLE C2: Two-way fixed-effects: Placebo tests with policy on in the pre- policy period 2008 -2012

	Dropout (1)	Dropout (2)	Dropout (3)
Repetition * Policy on 2008	0·024	0·019	0·002
	(0·018)	(0·017)	(0·015)
Repetition * Policy on 2009	−0·023	−0·026	−0·035*
	(0·022)	(0·020)	(0·19)
Repetition * Policy on 2010	−0·028	−0·028	−0·030
	(0·022)	(0·022)	(0·022)
Repetition * Policy on 2011	−0·020	−0·020	−0·017
	(0·019)	(0·019)	(0·019)
Repetition * Policy on 2012	−0·035	−0·036	−0·024
	(0·025)	(0·025)	(0·025)
Repetition * Policy on 2013	−0·144**	−0·159**	−0·148**
	(0·064)	(0·069)	(0·070)
Repetition * Policy on 2014	−0·356***	−0·367***	−0·347***
	(0·028)	(0·032)	(0·036)
School Fixed- Effects	Yes	Yes	Yes
Time Fixed-Effects	Yes	Yes	Yes
State * Year Fixed-Effects	No	Yes	No
County * Year Fixed-Effects	No	No	Yes
Other controls	Yes	Yes	Yes
R-squared	0·243	0·258	0·266
Number of schools	435822	435822	435822

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, clustered on state, in parenthesis. Regressions include interactions between grade repetition and year dummies taking the value of 0 in the pre-policy period and 1 in the subsequent years, 2008-2012 as if policy had changed before 2012. We use 2007 repetition rates as a base. Estimations include time, school fixed-effects and a combination of county-year fixed-effects. Regressions also include a set of school and teacher's characteristics as controls.

FIGURE C2: Average grade repetition 1st to 3rd and 4th to 6th grade: trends before and after policy. change



Bibliography

Adelman, N. E. et al. (1996): *The Uses of Time for Teaching and Learning.[Volume I: Findings and Conclusions.] Studies of Education Reform.*, ERIC.

Andrade-Baena, G. (2014): "Improving academic achievement through extended school-days: Evidence from escuelas de tiempo completo in mexico," Technical report, Masther Thesis. Standford University.

Angrist, J. and V. Lavy (1999): "Using maimonides' rule to estimate the effect of class size on scholastic achievement," *Quarterly Journal of Economics*, 114, 533–575.

Aronson, J., J. Zimmerman, and L. Carlos (1999): "Improving student achievement by extending school: Is it just a matter of time?." .

Babu, S. C. and J. A. Hallam (1989): "Socioeconomic impacts of school feeding programmes: empirical evidence from a south-Indian village," *Food Policy*, 14, 58–66.

Barro, R. and J.-W. Lee (2001): "International data on educational attainment: updates and implications," *Oxford Economic Papers*, 53, 541–563.

Bellei, C. (2009): "Does lengthening the school day increase students academic achievement? results from a natural experiment in chile," *Economics of Education Review*, 28, 629–640.

Bishop, J., W. Worner, and L. Weber (1988): "Extending the school day: An evaluation study of a seven-period class schedule," *Studies in Educational Evaluation*, 14, 361–380.

Brophy, J. (2006): "Grade repetition," *Education policy series*, 6, 420–437.

Brown, B. W. and D. H. Saks (1987): "The microeconomics of the allocation of teachers' time and student learning," *Economics of Education Review*, 6, 319 – 332.

Caldwell, J. H., W. G. Huitt, and A. O. Graeber (1982): "Time spent in learning: Implications from research," *Elementary School Journal*, 82, 471–480.

CARE (2004): "Annotated bibliography on food-assisted education," U.S Agency for International Development.

Cerdan-Infantes, P. and C. Vermeersch (2007): "More time is better: An evaluation of the full-time school program in uruguay." .

Chakraborty, T. and S. K. Bakshi (2016): “English language premium: Evidence from a policy experiment in india,” *Economics of Education Review*, 50, 1–16.

Coly, A. N., J. Milet, A. Diallo, T. Ndiaye, E. Bénéfice, F. Simondon, S. Wade, and K. B. Simondon (2006): “Preschool stunting, adolescent migration, catch-up growth, and adult height in young senegalese men and women of rural origin,” *The Journal of nutrition*, 136, 2412–2420.

CONAPO (2010): “Indice de marginacion por localidad.” Technical report, Consejo Nacional de Poblacion.

CONEVAL (2013): “Informe de la evaluación específica de desempeño PETC 2012-2013,” Technical report, Consejo Nacional de Evaluacion de la Politica de Desarrollo Social (CONEVAL.

Cuban, L. (2008): “The perennial reform: Fixing school time.” *Phi Delta Kappan*, 90, 240–250.

Cueto, S. and M. Chinen (2008): “Educational impact of a school breakfast programme in rural Peru,” *International Journal of educational development*, 28, 132–148.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006): “Interpreting the evidence on life cycle skill formation,” *Handbook of the Economics of Education*, 1, 697–812.

Del Rosso, J. M. (1996): *Class Action: Improving School Performance in the Developing World Through Better Health and Nutrition. Directions in Development*, World Bank-free PDF.

DIF (2011): “Evaluacion interna: Programa de desayunos escolares Distrito Federal,” .

Dong, Y. (2010): “Kept back to get ahead? kindergarten retention and academic performance,” *European Economic Review*, 54, 219–236.

Duflo, E. (2000): “Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment,” Technical report, National Bureau of Economic Research.

Eveleth, P. B. and J. J. M. Tanner (1976): *World-wide variation in human growth*, volume 8, CUP Archive.

Faguet, J.-P. and F. Sánchez (2008): “Decentralizations effects on educational outcomes in bolivia and colombia,” *World Development*, 36, 1294 – 1316.

Galiani, S., P. Gertler, and E. Schargrodsky (2008): "School decentralization: Helping the good get better, but leaving the poor behind," *Journal of Public Economics*, 92, 2106–2120.

Glewwe, P. W., E. A. Hanushek, S. D. Humpage, and R. Ravina (2011): "School resources and educational outcomes in developing countries: a review of the literature from 1990 to 2010," Technical report, National Bureau of Economic Research.

Gomes-Neto, J. B. and E. A. Hanushek (1994): "Causes and consequences of grade repetition: Evidence from brazil," *Economic Development and Cultural Change*, 117–148.

Gómez, M., M. Flores, and L. Alemán (2013): "The full-time school program in Mexico," *Journal of Case Studies in Education*.

Grantham-McGregor, S. (2005): "Can the provision of breakfast benefit school performance?" *Food & Nutrition Bulletin*, 26, 144–158.

Grantham-McGregor, S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, and B. Strupp (2007): "Developmental potential in the first 5 years for children in developing countries," *The Lancet*, 369, 60–70.

Gundersen, C., B. Kreider, and J. Pepper (2012): "The impact of the national school lunch program on child health: A nonparametric bounds analysis," *Journal of Econometrics*, 166, 79–91.

Gutierrez, J., J. Rivera, and T. Shamah (2012): "National survey of nutrition and health 2012," National Institute of Public Health, cuernavaca Morelos, Mexico.

Hanushek, E. and D. Kimko (2000): "Schooling, labor-force quality, and the growth of nations," *American economic review*, 1184–1208.

Hanushek, E. A. (2003): "The failure of input-based schooling policies*," *The economic journal*, 113, F64–F98.

Hanushek, E. A. (2004): "Some simple analytics of school quality," Technical report, National Bureau of Economic Research.

Hanushek, E. A. (2013): "Economic growth in developing countries: The role of human capital," *Economics of Education Review*, 37, 204–212.

Heckman, J. J. (2008): "Schools, skills, and synapses," *Economic inquiry*, 46, 289–324.

Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an econometric evaluation estimator," *The Review of Economic Studies*, 65, 261–294.

Heckman, J. J., R. Pinto, and P. A. Savelyev (2012): "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," Technical report, National Bureau of Economic Research.

Heckman, J. J. and R. Robb Jr (1985): "Alternative methods for evaluating the impact of interventions: An overview," *journal of Econometrics*, 30, 239–267.

Heinrich, C., A. Maffioli, and G. Vazquez (2010): "A primer for applying propensity-score matching," Technical report, Inter-American Development Bank, Office of Strategic Planning and Development Effectiveness (SPD).

Hinrichs, P. (2010): "The effects of the National School Lunch Program on education and health," *Journal of Policy Analysis and Management*, 29, 479–505.

Hirvonen, K. (2013): "Measuring catch-up growth in malnourished populations," *Annals of Human Biology*.

Holland, P. A., P. Alfaro, and D. Evans (2015): "Extending the school day in latin america and the caribbean," *World Bank Policy Research Working Paper*.

Holmes, C. T. et al. (1989): "Grade level retention effects: A meta-analysis of research studies," *Flunking grades: Research and policies on retention*, 16, 33.

Hussain, I. (2012): "Subjective performance evaluation in the public sector: evidence from school inspections," *Unpublished Manuscript. University of Sussex*.

Jacob, B. A. and L. Lefgren (2004): "Remedial education and student achievement: A regression-discontinuity analysis," *Review of economics and statistics*, 86, 226–244.

James-Burdumy, S., M. Dynarski, M. Moore, J. Deke, W. Mansfield, C. Pistorino, and E. Warner (2005): "When schools stay open late: The national evaluation of the 21st century community learning centers program. final report." *US Department of Education*.

Jimerson, S. R., G. E. Anderson, and A. D. Whipple (2002): "Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school," *Psychology in the Schools*, 39, 441–457.

Karweit, N. (1985): "Should we lengthen the school term?" *Educational Researcher*, 14, 9–15.

Koppensteiner, M. F. (2014): “Automatic grade promotion and student performance: Evidence from brazil,” *Journal of Development Economics*, 107, 277–290.

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

Levin, H. M. and M. C. Tsang (1987): “The economics of student time,” *Economics of Education Review*, 6, 357 – 364.

Link, C. R. and J. G. Mulligan (1986): “The merits of a longer school day,” *Economics of Education Review*, 5, 373–381.

Lochner, L. (2011): “Non-production benefits of education: Crime, health, and good citizenship,” Technical report, National Bureau of Economic Research.

Machin, S. and S. McNally (2008): “The literacy hour,” *Journal of Public Economics*, 92, 1441–1462.

Manacorda, M. (2012): “The cost of grade retention,” *Review of Economics and Statistics*, 94, 596–606.

Meyer, E. and C. Van Klaveren (2013): “The effectiveness of extended day programs: Evidence from a randomized field experiment in the netherlands,” *Economics of Education Review*, 36, 1–11.

Millimet, D. L. and R. Tchernis (2012): “Estimation of treatment effects without an exclusion restriction: with an application to the analysis of the school breakfast program,” *Journal of Applied Econometrics*.

Millimet, D. L., R. Tchernis, and M. Husain (2010): “School nutrition programs and the incidence of childhood obesity,” *Journal of Human Resources*, 45, 640–654.

Murnane, R. J., J. B. Willett, Y. Duhaldeborde, and J. H. Tyler (2000): “How important are the cognitive skills of teenagers in predicting subsequent earnings?” *Journal of Policy Analysis and Management*, 19, 547–568.

Murnane, R. J., J. B. Willett, and F. Levy (1995): “The growing importance of cognitive skills in wage determination,” Technical report, National Bureau of Economic Research.

NECTL (2005): “Prisioners of time: too much to teach, not enough time to teach it,” Technical report, National Education Commission on Time and Learning.

Patall, E. A., H. Cooper, and A. B. Allen (2010): "Extending the school day or school year a systematic review of research (1985–2009)," *Review of educational research*, 80, 401–436.

Patrinos, H. A., L. F. López-Calva, G. Bando, et al. (2005): "Child labor, school attendance, and indigenous households: evidence from mexico," *World Bank Policy Research Working Paper*.

Pollitt, E. (1995): "Does breakfast make a difference in school?" *Journal of the American Dietetic Association*, 95, 1134–1139.

Pollitt, E., S. Cueto, and E. Jacoby (1998): "Fasting and cognition in well-and undernourished schoolchildren: a review of three experimental studies." *The American journal of clinical nutrition*, 67, 779S–784S.

Powell, C., S. Walker, S. Chang, and S. Grantham-McGregor (1998): "Nutrition and education: a randomized trial of the effects of breakfast in rural primary school children." *The American journal of clinical nutrition*, 68, 873–879.

Prentice, A. M., K. A. Ward, G. R. Goldberg, L. M. Jarjou, S. E. Moore, A. J. Fulford, and A. Prentice (2013): "Critical windows for nutritional interventions against stunting," *The American journal of clinical nutrition*, 97, 911–918.

Protheroe, N. (2007): "Alternatives to retention in grade," *Principal, January/February*, 30–34.

Ramirez, E., M. Grijalva, M. Valencia, J. Ponce, and E. Artalejo (2005): "Effect of a school breakfast program on the prevalence of obesity and cardiovascular risk factors on children," *Salud Publica de Mexico*, 47.

Raven, J. (1998): "Manual for ravens progressive matrices and vocabulary scales," *Oxford: Oxford Psychologists*, 12, G60p.

Robin, K. B. (2005): *The effects of extended-day, extended-year preschool on learning in literacy and mathematics*, ProQuest.

Roderick, M. (1994): "Grade retention and school dropout: Investigating the association," *American Educational Research Journal*, 31, 729–759.

Rodríguez, E. (2005): *Sin herencia de pobreza: el programa Progresa-Oportunidades de México*, Idb.

Schanzenbach, D. W. (2009): “Do school lunches contribute to childhood obesity?” *Journal of Human Resources*, 44, 684–709.

SEP (2010): “Orientaciones pedagógicas para la organización del trabajo en las escuelas de tiempo completo.” Technical report, Dirección General de Desarrollo Curricular. Subsecretaría de Educación Básica.

Shamah, T., M. Morales, and A. Hernandez (2010): “Contribution of the school breakfast on the nutritional status in the state of mexico (in Spanish),” National Institute of Public Health, cuernavaca Morelos, Mexico.

Shemilt, I., I. Harvey, L. Shepstone, L. Swift, C. Reading, M. Mugford, P. Belderson, N. Norris, J. Thoburn, and J. Robinson (2004): “A national evaluation of school breakfast clubs: evidence from a cluster randomized controlled trial and an observational analysis,” *Child: care, health and development*, 30, 413–427.

Skoufias, E. and J. Shapiro (2006): “Evaluating the impact of mexico’s quality schools program: The pitfalls of using nonexperimental data,” .

Slattery, P. (1995): “A postmodern vision of time and learning: A response to the national education commission report prisoners of time,” *Harvard Educational Review*, 65, 612–634.

SNDIF (2012): “White book. integral strategy for social and alimentary assistance (EIASA): Particulay the breakfast program,” .

Soto, J. and R. Lorenzo (2008): “Reporte de campo: Visitas a cuadro sedif,” Centro de Estudios Estrategicos: Tecnologico de Monterrey, mexico, D.F.

Stuart, E. A. (2010): “Matching methods for causal inference: A review and a look forward,” *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25, 1.

UNESCO (2010): “Estudio de caracterización de las escuelas de tiempo completo.” Technical report.

Vera, J. A. (2005): “Evaluation of the impact that school breakfasts have on cognitive-motor development (in Spanish),” *Psicologia y Ciencia Social*, 5.

Victora, C. G., M. de Onis, P. C. Hallal, M. Blössner, and R. Shrimpton (2010): “Worldwide timing of growth faltering: revisiting implications for interventions,” *Pediatrics*, 125, e473–e480.

Wheeler, P. (1987): "The relationship between grade six test scores and the length of the school day." *Educational Research Quarterly*, 11, 10–17.