University of Sussex
Economics

Working Paper Series

No. 74-2015

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

Francisco Cabrera-Hernandez
Department of Economics, University of Sussex
F.Cabrera-Hernandez@sussex.ac.uk

Abstract: Mexican educational authorities face a significant and challenging problem of low achievement in standardized tests applied to pupils in primary schools. This research looks at a Full-Time Primary Schools Program implemented in 2007, to work out if changing the time pupils spend at school and a modification in the structure of teaching can enhance skills in language and mathematics. The results of Differences in Differences (DiD) and Propensity Score Match plus DiD, point to a significant impact of the program with an improvement of 0.11 SD on mathematics and Spanish test scores after four years of treatment. More importantly, these improvements are significantly higher in schools located in deprived areas, ranging from 0.12 SD to 0.29 SD on both subjects after two and four years of treatment, respectively. The impacts also show a significant average decrease in the proportion of students graded as 'insufficient', combined with an increase of those graded as 'excellent'. Further analysis on causal channels shows that policy effects do not come from changes in the composition of pupils in treated schools. These findings are of strong significance when placed into the wider education debate about what works best in schools for improving pupil performance.

JEL classification: I2, I21

Keywords: full-time schools, test scores, school reform, time of instruction, school's inputs

Acknowledgements: I want to thank Richard Dickens, Shqiponja Telhaj and Iftikhar Hussain for their valuable comments and suggestions about the design and results of this evaluation. I am also very thankful to the specialists of the Secretariat of Public Education (SEP) who helped me understand policy details and implementation during the workshop: "Incubadora de Evaluaciones de Impacto" organized by CONEVAL and J-pal in Mexico city. I am responsible for all remaining errors.

1 Introduction

According to the 2009 results of the Programme for International Student Assessment (PISA), 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.

Economic research shows that improvements on math, language and science test scores relate to increases in real annual 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). There are also other non-monetary benefits from education such as improved health status and lowered crime.[1] However these studies do not provide a clear guidance of what policies and specific investments should be pursued to increase educational outcomes.

Several policies directed to schools have shown to raise enrollment, however, experimental and non-experimental research have not shown bold evidence of large effects on learning from diverse public interventions. 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.[3]

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[4] 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 dis-

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

[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] For a detailed review of the evidence of the effect of different school policies on educational outcomes in developing countries, see Glewwe, Hanushek, Humpage, and Ravina (2011)

[4] 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)

cussed in the United States (US) (Link and Mulligan, 1986, Levin and Tsang, 1987, Brown and Saks, 1987, Slattery, 1995, NECTL, 2005). 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 Excuelas de Tiempo Completo*, or PETC) on primary academic achievement,[5] measured in standardized test scores of mathematics and Spanish from the 2008-2009 to the 2012-2013 academic year.[6] 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.[7] 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 millions 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.

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

[5]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.

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

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

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 this 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;[8] 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, the strategy is refined by the computation of a PSM that pairs similar schools between the original treated and control groups and hence new DiD estimations are obtained.

Estimations 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 aapproximately 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 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. 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

[8] Specific characteristics of this test will be discussed in detail in Section 3

targeting of the program in Mexico and for its implementation in other developing countries.

The rest of this study is presented as follows. Section 2 discusses prior evidence on full-time school programs. Section 3 outlines the main characteristics of PETC since its inception. Section 4 presents the data and includes descriptive statistics. Section 5 discusses the empirical strategy and presents the main results. Section 6 discusses some of the impact channels of PETC on test scores. Section 7 concludes.

2 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) [9]

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 in the US by randomly assigning 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

[9] 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)

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

Meyer and Van Klaveren (2011) 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 neither 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.

3 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).[10]

Although the curricula for PETC schools is flexible, the program allows for a specific time (i.e. one

[10]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)

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.[11]

For the purposes of the program, schools should preferably have a dinning 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).[12]

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.

3.1 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.[13]

[11] An of example of the timetable suggested for PETC schools can be found in Table A1in the Appendix.

[12] 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 with about 100 students, 5 teachers and 1 principal, in 2014. The budget for this school was approximately US$40,000 of which around US$14,000 are fixed. 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.

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

- 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 dinning rooms).
- 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.[14] 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).

3.2 ENLACE Test and patterns of application

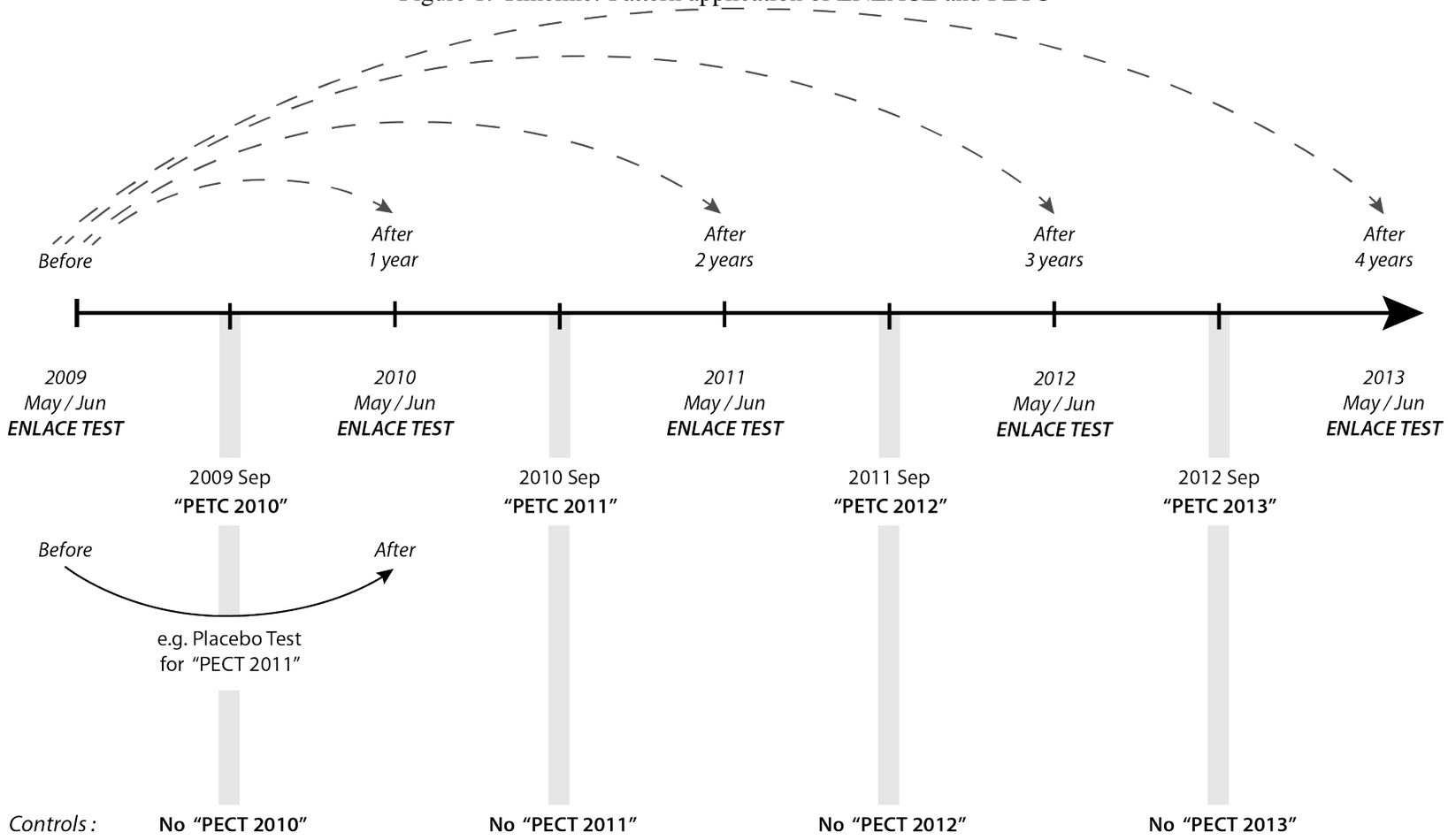
ENLACE is a census standardized exam of mathematics and Spanish (plus one extra subject, i.e. science or history rotating every year) directed to evaluate knowledge and skills of students from third to sixth grade of primary education, first to third grade of secondary education and first year of high 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.

[14] 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 4.

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 1, the data at hand allows 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 PECT schools started the program in 2011, 2012 and 2013.

Figure 1: Timeline: Pattern application of ENLACE and PETC



3.3 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, Ichimura, and Todd (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.[15]

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 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 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.

Table 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 then selected for the difference-in-difference analysis.

[15]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 A7 to A14 in the Appendix.

4 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.[16] 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.[17]

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, 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.

4.1 Descriptive statistics

Table 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

[16] 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.

[17] 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)

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. Also, unobserved heterogeneity is accounted in the trends of results presented by treated and controls schools as discussed in detail in section 5.

...

Table 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 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 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 3: Mean outcomes for various samples

		Standardized Test Scores				
	Number of Schools	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

5 Impact of PETC on Test Scores and Grade Repetition

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.[18] 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. 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:

$$Y_{st} = \beta PETC_s + \gamma t_s + \delta_1 (PETC_s * t_s) + \delta_2 X_{st} + \pi_e + \mu_{st} \tag{1}$$

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 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.[19]

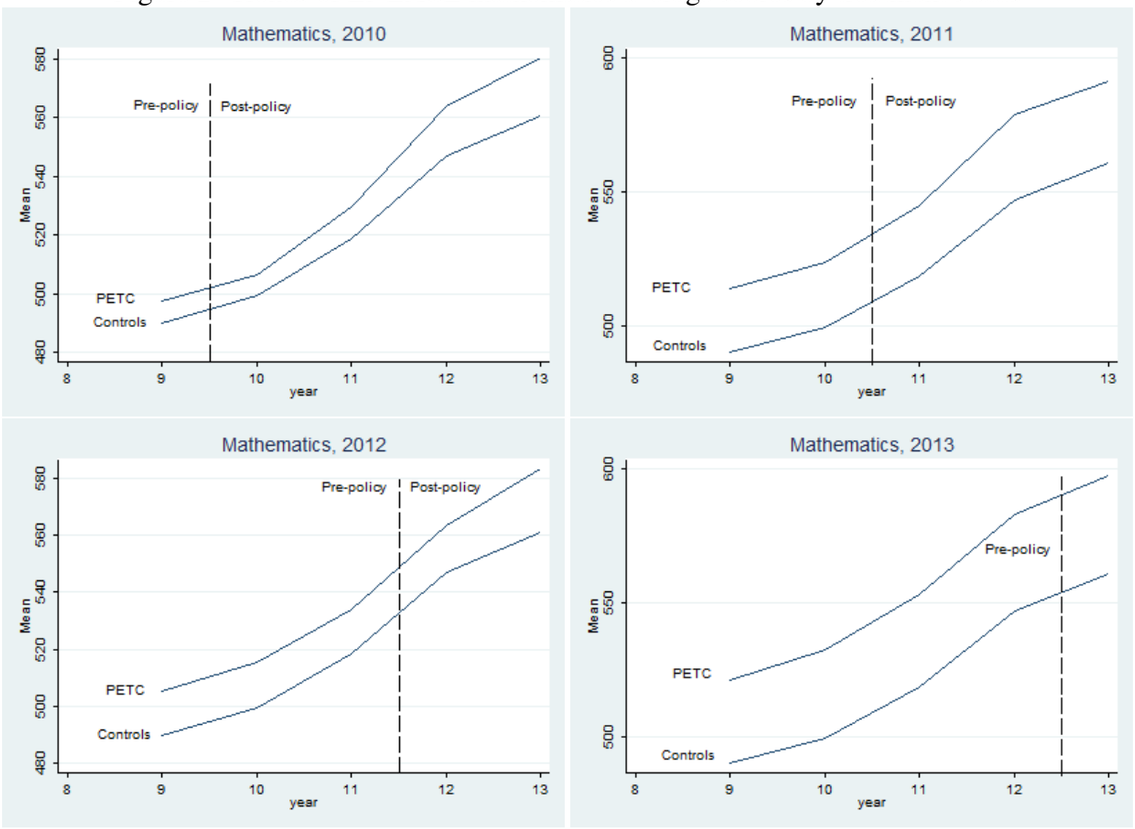
Next subsection includes DiD and PSM plus DiD estimations for pre-policy period (i.e. placebo tests)

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

[19] 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 A1, A2 and A3 in the Appendix.

to discard significant unobserved differences between treated and control schools in the base period, once a rich set of control variables are included.

Figure 2: Trends of ENLACE mathematics average scores by treatment status



5.1 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 4 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 4 only suggest a possible threat for the conclusions of the effects on Spanish test scores, specifically for schools treated in 2013.

Table 4: 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)	(0.012)	(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.

5.2 Basic DiD results

Table 5 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 5: 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 6 shows the results for the effects of PETC on Spanish. It presents the same specifications than Table 5. 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 6: 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.

5.3 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 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.[20] 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.[21]

Table 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	90586	73939	82518	82007
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	90585	73935	82513	82007
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 3. It can be observed that altough the effect on low

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

[21] All PETC effects on math test scores separated by cohort and years of treatment can be found in the Appendix Table A2 using all non-PETC schools and in Table A3 using a matched control group. For the case of Spanish these can be found in the Appendix Table A4 and Table A5, respectively.

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.

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

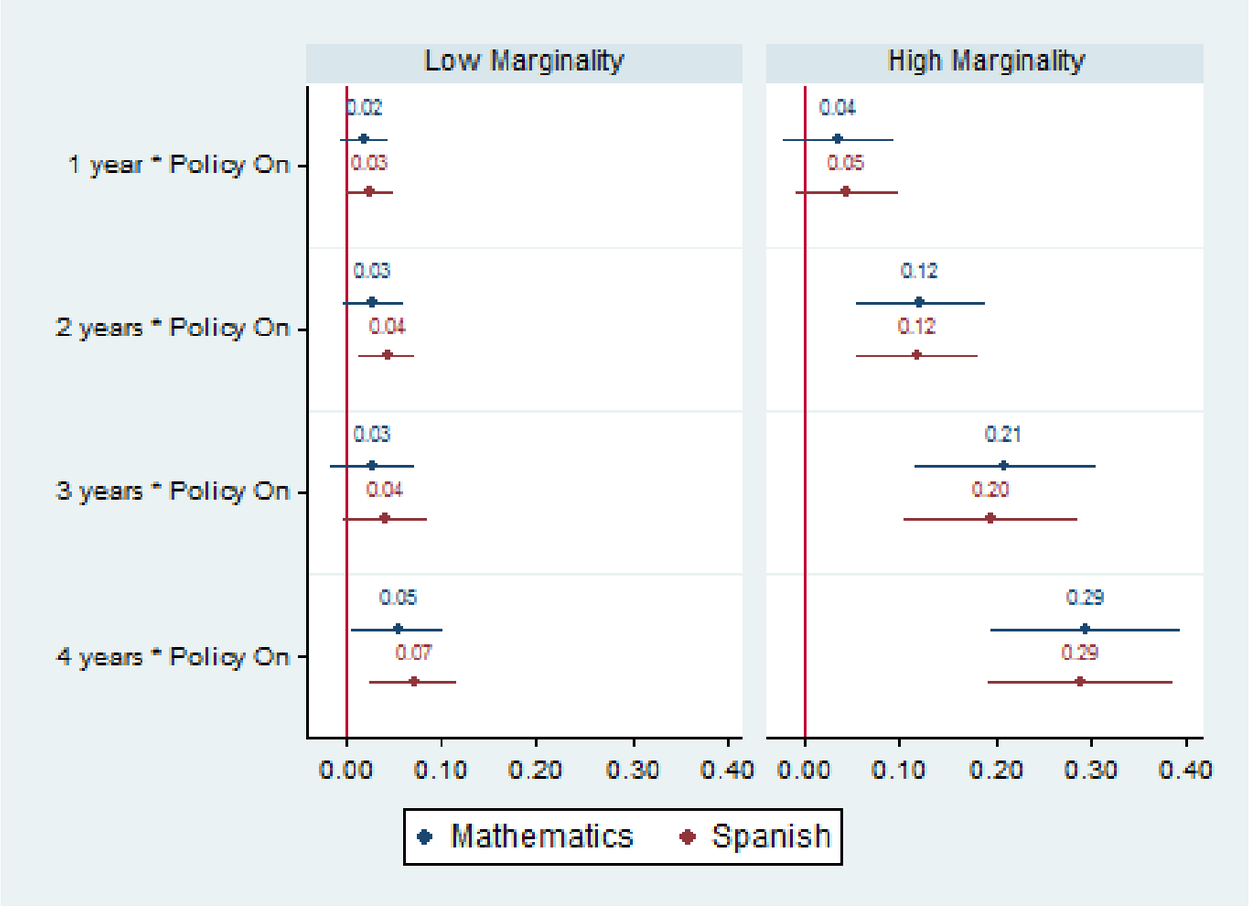


Figure obtained from the point estimators and the 95% confidence intervals coming from a DID regression including a set of dummy variables interacting a post policy dummy with the number of years since intervention. The regression also includes school fixed effects, time-fixed effects and full set of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

Figure 4 shows heterogenous 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 schores but this difference reduces and practically dissapears 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.

Figure 4: Average effects of PETC on mathematics and Spanish standardized test scores by PEC participation and years of treatment

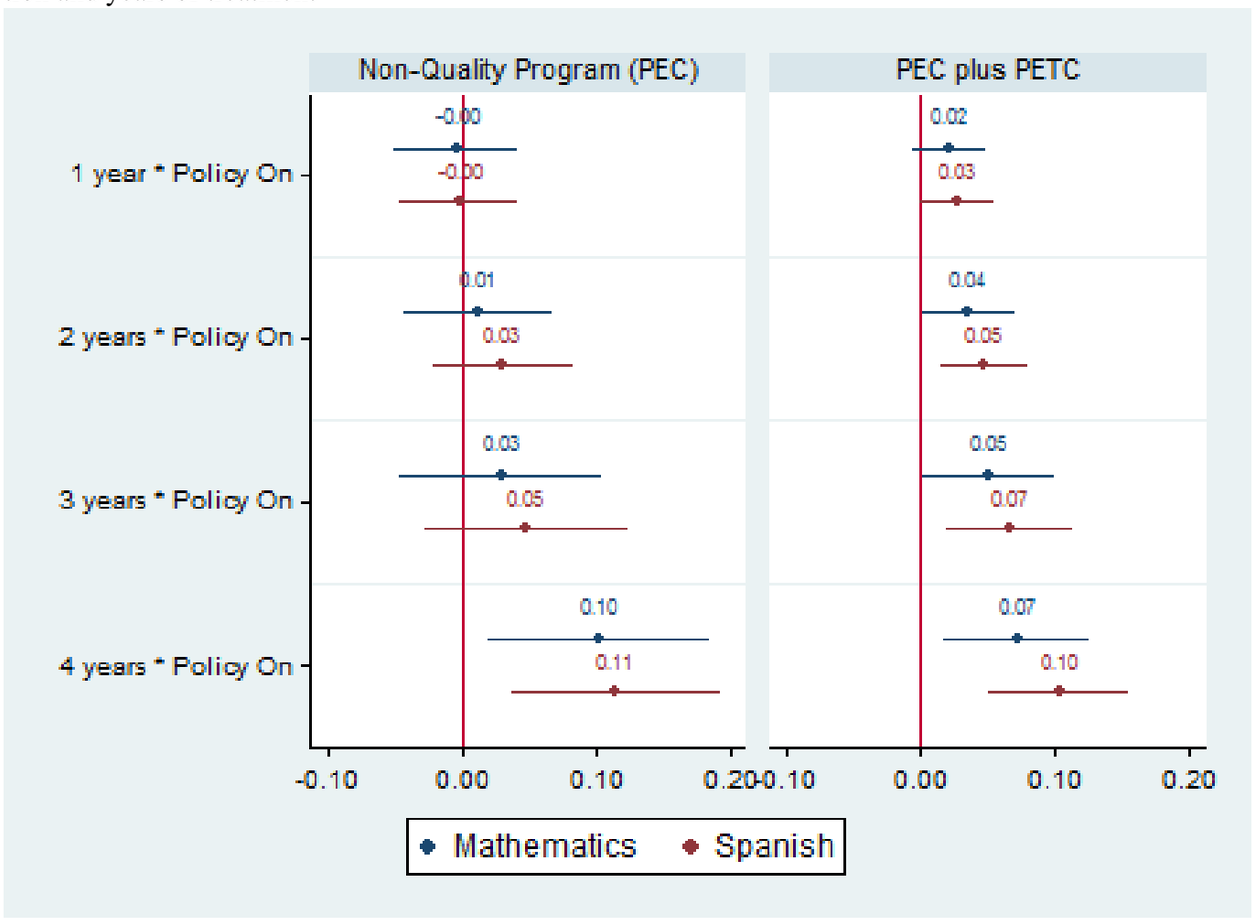


Figure obtained from the point estimators and the 95% confidence intervals coming from a DID regression including a set of dummy variables interacting a post policy dummy with the number of years since intervention. The regression also includes school fixed effects, time-fixed effects and full set of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

6 Impact channels of the effects of PETC on test scores

6.1 Is PETC having an effect on students with different abilities?

Table 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 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 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%).

Results for Spanish are presented in Figure 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,

Figure 5: Average effects of PETC on the distribution of mathematics standardized test cores by marginality level and years of treatment

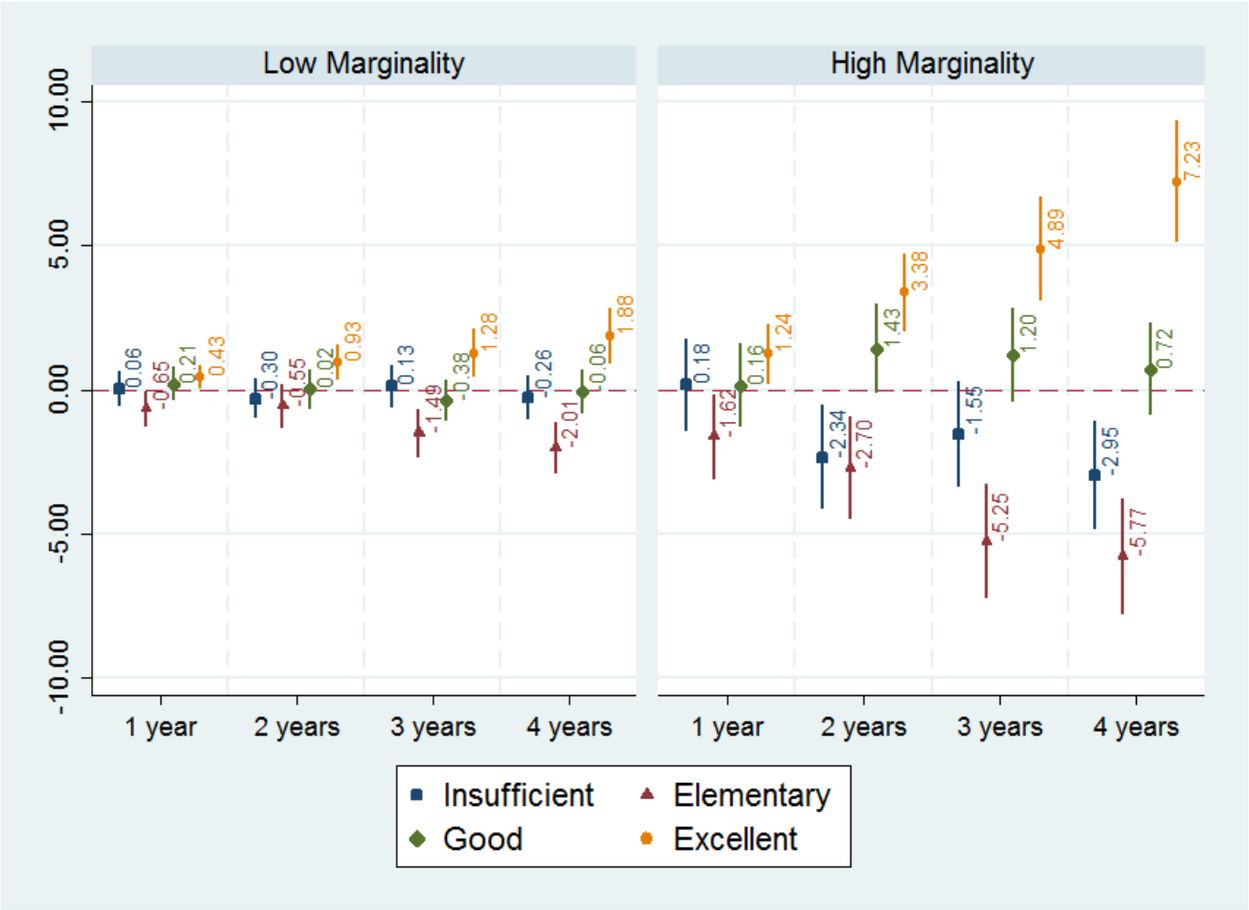


Figure obtained from the point estimators and the 95% confidence intervals coming from a DID regression including a set of dummy variables interacting a post policy dummy with the number of years since intervention. The regression also includes school fixed effects, time-fixed effects and full set of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

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%.

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

Figure 6: Average effects of PETC on the distribution of Spanish standardized test cores by marginality level and years of treatment

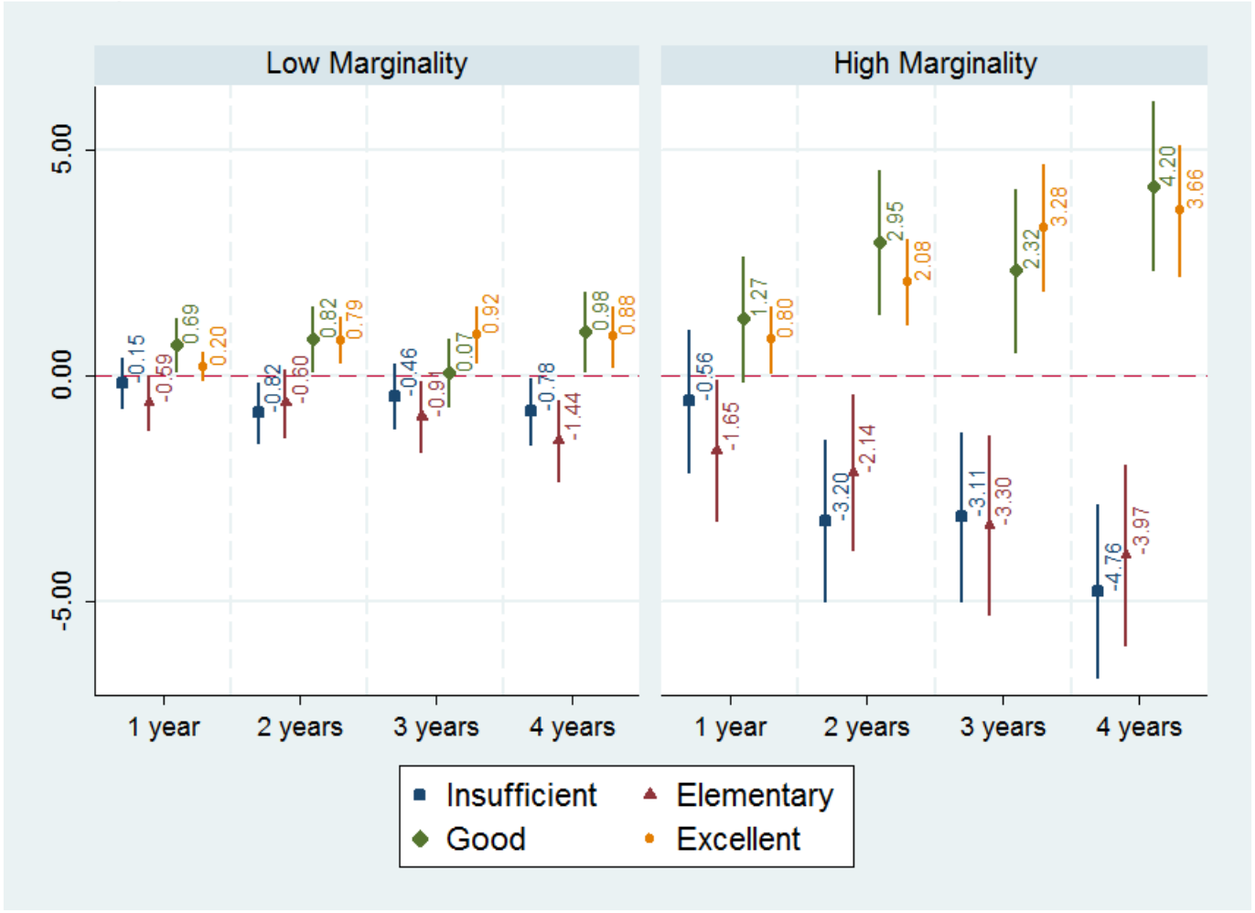


Figure obtained from the point estimators and the 95% confidence intervals coming from a DID regression including a set of dummy variables interacting a post policy dummy with the number of years since intervention. The regression also includes school fixed effects, time-fixed effects and full set of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

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 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.[22] 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.[23]

6.2 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

[22] Schools that present a positive inflow of students are analyzed separately below.
[23] Placebo tests on dropout rates are presented in Table A6 in the Appendix

Table 9: Effects of PETC on dropout rates by level of marginality

Control Schools	(1)	(2)	(3)	(4)	(5)	(6)
	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

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

Columns 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.

best new students, who would on average present better results in standardized tests.[24] Consequently, the positive results of the program as discussed before, may well be explained by selection rather than policy intervention.

The results shown so far may well contend 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 in the lower parts of the distribution of scores. 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 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.

[24] 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.

Table 10: Effects of the proportion of new students at any given grade and year in PETC Schools on Spanish and mathematics test scores

	(1)	(2)	(3)
	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
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

7 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 showing that there are no trend differences in 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.

References

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?." .

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.

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.

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

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.

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.

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

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

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

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

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*.

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

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.

Meyer, E. and C. Van Klaveren (2011): "Evaluation of an extended day program in the netherlands: a randomized field experiment," Technical report.

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.

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

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.

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.

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

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

Appendices

Table A1: 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 A1: Trends of Spanish average scores ENLACE by treatment status

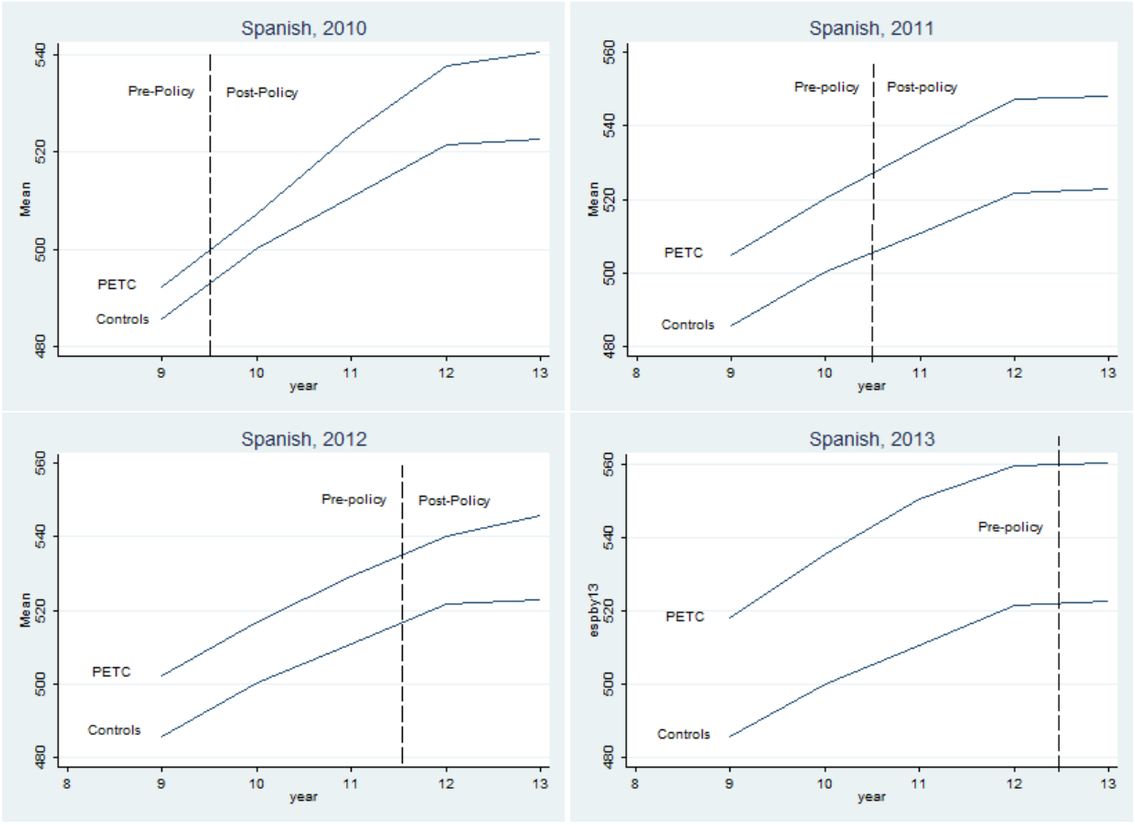


Figure A2: Trends of math average scores ENLACE after PSM by treatment status

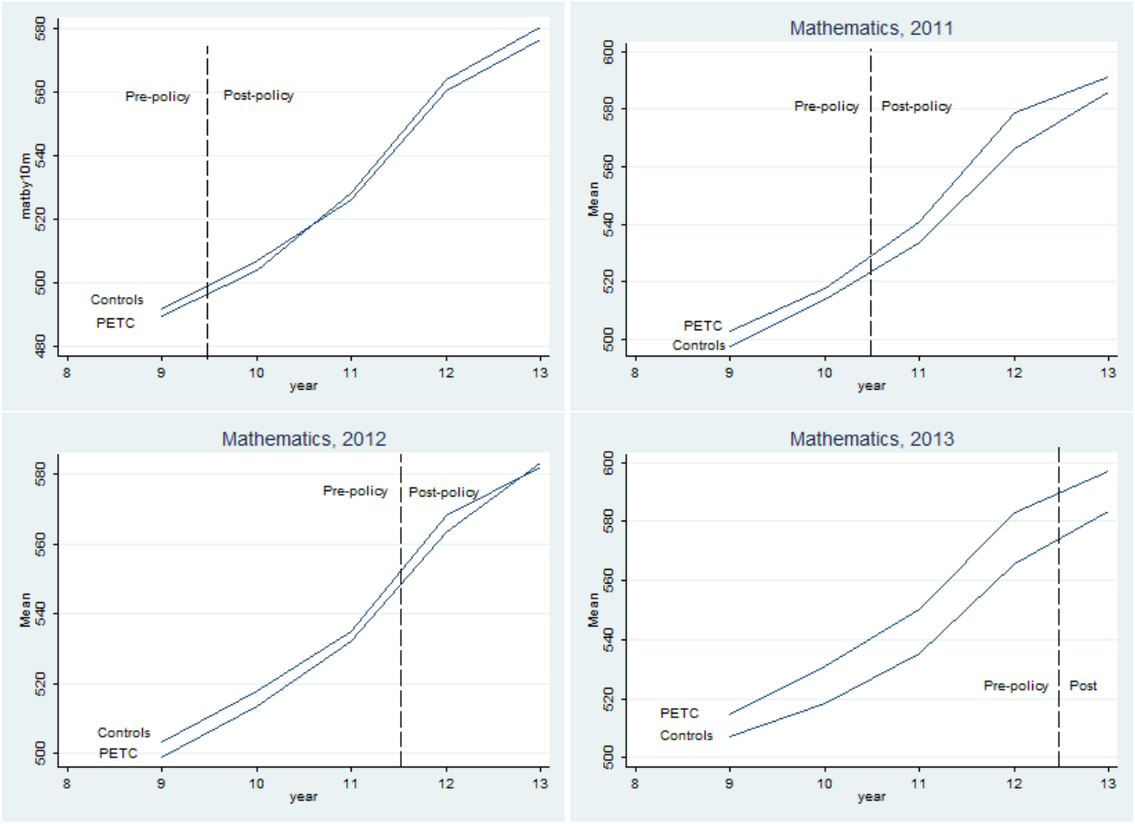


Table A2: Differences in Differences: Standardized Mathematics Test Scores by PETC Cohort

	(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 A3: 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 A4: 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 A5: 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 A3: Trends of Spanish average scores ENLACE after PSM by treatment status

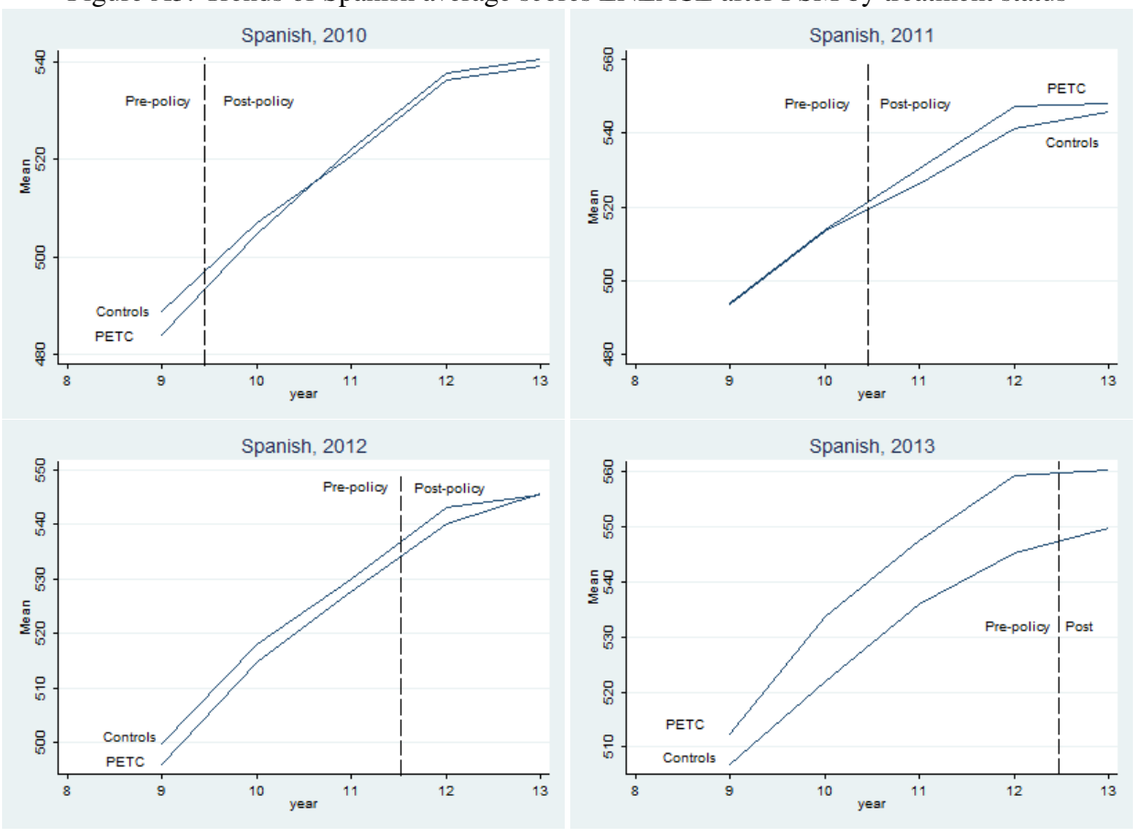


Table A6: Placebo regressions: DiD and PSM-DiD for dropout rates

	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 A7: 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 A8: 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 A9: 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 A10: 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 A11: 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 A12: 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 A13: 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 A14: 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