Undergraduate Economic Review

Volume 16 | Issue 1 Article 8

2019

Learning Consequences of School Improvement in Mexico: Evidence from a Large Government Program

Carlos Alejandro Noyola Contreras
University of Bristol, carlos_no_yola@hotmail.com

Learning Consequences of School Improvement in Mexico: Evidence from a Large Government Program

Abstract

I study the impact of investment in infrastructure of already existing poor schools and increased school based management on learning outcomes, as measured by student achievement in standardized tests. To that end, I implement a difference-in-differences design to compare schools that received money from a large government program to improve their physical conditions with those that do not, before and after program implementation. Unlike previous studies, I focus on the effect of improving schools that already exist, to see whether the impact is different from that of building schools. I find no evidence of positive impacts on test scores at the school level, and some evidence of a negative impact for secondary schools.

Keywords

Education, Infrastructure, Government program, Mexico, Learning outcomes, standardized test scores

Cover Page Footnote

Thanks to Cristián Sánchez for his support.

1 Introduction

This paper studies the impact of investment in infrastructure of already existing poor schools and increased school based management on learning outcomes, as measured by student achievement in standardized tests. In 2014, the Mexican government implemented the Program for Schools of Excellence (PEE in Spanish), aimed at helping poor schools improve their infrastructure and strengthen their management autonomy. In that year, with a budget of 7,600 million pesos, the program benefited 20,000 schools (kindergarden, elementary and junior high schools), and it was the first of its kind in Mexico to assign money directly to schools at a large scale. I use a difference in differences design to compare the percentage of students in the two highest levels of attainment of schools that benefited from the program with those that did not, before and after program implementation. I find no evidence of positive impacts on test scores, and some evidence that the program decreased the percentage of students in the highest levels of attainment.

This paper contributes to the literature in three ways. First, most of the research studying the effects of investments in infrastructure is focused on new schools. Improved infrastructure of already existing schools has received little attention, in part because there have not been many large scale programs improving infrastructure. However, as the results from one program in Bolivia suggest, the effect of rehabilitation might not be the same as that of building new schools. Thus, it is important to understand how do effects from improving facilities differ from those of building new schools. Second, PEE was one of the main government programs implemented in Mexico in the last few years as part of the Educational Reform, aimed at substantially improving the quality of education throughout the country. Therefore, it is crucial for policy implications to evaluate the program in terms of learning outcomes in order to understand whether or not it achieved its goal. Even though 300 schools benefited were evaluated one year after the implementation, the sample, according to the authors, was not chosen in a way that it is representative, and the evaluation was mainly based on surveys conducted with principals, teachers and parents about their perception of the program and its outcomes (Valora Consultoría S.C., 2015). Thus, it remains to evaluate the program in terms of the impact on measurable outcomes of students. Finally, it contributes to put into question the results in the literature found by Glewwe, et.al, and McEwan about the impact of infrastructure and resources on learning, also possibly with policy implications.

Central to economic development is understanding how can we improve education in underdeveloped countries. Although there is a growing literature in the field, it is not clear that simply by increasing investments education will improve. For instance, Hanushek and Woessmann found that educational expenditure per student increased substantially in real terms in many OECD countries between 1970 and 1990, but none of them has had significant improvements in average student achievement, measured by results on PISA tests (Hanushek and Woessmann, 2011). On the other hand, others have found that increases in per pupil spending in a given district increase not only the number of school years completed, but also earnings in adulthood and reduce the probability of being poor (Jackson, Johnson and Persico, 2015). Therefore, it is key to understand what types of investments have positive impacts on education.

One of the basic concerns about education expenditures is whether investments in infrastructure and school management improve educational outcomes (Duflo, 2001). Some studies found that by reducing travel time, building new schools significantly increased enrollment rates in Afghanistan and Pakistan, as well as test scores in the case of the former, although most of the increase in test scores was due to the fact that most children were not enrolled at school before the program (Krishnaratne, White and Carpenter, 2013). In another study, Duflo studied the effects of a government program in Indonesia that constructed new schools in regions where enrollment rates were the lowest, and found evidence that it increased the number of years of education and wages in adulthood (Duflo, 2001). Results of the BRIGHT program in Burkina Faso (that constructed new schools first and then improved them) also found supportive evidence of increased enrollment and test scores as a result of the program (Krishnaratne, White and Carpenter, 2013).

These results suggest that investments in infrastructure improve education. However, building new schools and improving the infrastructure of schools already established is not the same thing. Newman and his colleagues studied the effects of small scale rural infrastructure projects in education in Bolivia, and found little impact on education outcomes (Newman, et. al., 2002). Glewwe, et.al, studied 79 papers of the effect of different types of investment on education, and found that while there is strong evidence supporting the idea that some investments in infrastructure (desks, tables, chairs, blackboards, better walls, roofs and floors) improve learning outcomes, evidence for improvements in electricity and other electronic resources, such as computers, is weak at best (Glewwe, et.al, 2011). McEwan gathered 77 randomized experiments that evaluate the impacts of school interventions on learning in developing countries and found the largest average effect sizes to be the ones for treatments that incorporate instructional materials, computers or instructional technology (McEwan, 2015). However, McEwan also found that instructional material by itself does not improve outcomes, rather it is a complement for teacher training and a well articulated instructional model (McEwan,2015).

As for management, McEwan found the effects of treatments that improve school management to be very small, such as more school-based management policies (McEwan,2015). In contrast, a program in Kenya that trained local school committees to evaluate the performance of teachers and decide whether or not to renew their contracts was found to have significant positive impacts on school enrollment (Krishnaratne, White and Caperpenter, 2013). Finally, a study of a program implemented in Mexico in the last decade that decentralized management decisions to the school level found that it reduced dropout, failure and repetition rates (Skoufias and Shapiro, 2006).

2 Education in Mexico

Quality of public education in Mexico is poor. According to the Programme for International Student Assessment (PISA) in 2015, Mexico ranks the lowest among OECD countries, and it has been like that in the past 15 years. Moreover, Mexican students, on average, fail every subject evaluated by PISA: mathematics, science and reading.

According to INEGI, the National Statistics Institute, the illiteracy rate was above 10 per cent in 8 states in 2010, but only 4 had rates above 10 per cent in 2015. Nationwide, for example, the illiteracy rate dropped from more than 10 per cent in 1990 to less than 7 per cent in 2010. Similarly, in 2010 more than half of the states had an average length of schooling below 9 years, that is, not even secondary school, and some even below 7 years. Five years later none of the states had an average length of schooling below 7 years, and in more than half of the states the average person aged 15 or older had completed at least 9 years of school.

In other words, Mexican students are spending more time at the classroom without learning. Given that Mexico spends more on education, measured as a percentage of GDP, than developed countries like Germany and Spain (5.2 per cent in 2012, according to OECD), one of the challenges is to translate time spent at school into better learning outcomes.

The PEE program is part of the Educational Reform, approved by congress in 2012 as a major effort to improve quality of education. Among others, the Reform included the creation of the Professional Teacher Service (Servicio Profesional Docente), a new set of rules according to which teachers would be evaluated to determine whether further training was needed, as well as eligibility of aspiring teachers.

The program had a budget of 7,600 million pesos, and although benefited 20,000 schools (kindergarden, elementary and junior high schools), I only use elementary and junior high schools, since those are the only ones subject to evaluation using standardized tests. To select the schools, the government created an index of infrastructural poverty (ICE in Spanish) to divide schools in 5 categories: very low, low, medium, high and very high. The index is continuous and takes values from 0 to 1, where 0 is no infrastructural poverty and 1 means complete poverty. Five variables are taken into account to evaluate poverty of the school: type of building, building material, water availability, bathroom availability and basic equipment at the classroom, where basic equipment takes into account whether all classrooms have blackboards, whether all students have a table and a chair, and whether all teachers have a table and a chair.

Only schools classified with high or very high infrastructural poverty were eligible to receive aid from the program. After that, schools had to comply with the requirements that the law specifies for them to receive money from a government program. The government claimed to select 20,000 schools from the two highest levels of the index after evaluating their suitability (Valora Consultoría S.C., 2015). Then, the amount of money was assigned according to the number of students. The money assigned to each school was divided in two: money for the improvement of physical infrastructure and money devoted to strengthen the autonomy of the school (solving problems of operation, preventing school dropouts and strengthening reading, writing and math skills of students). The schools had to inform of the use of the resources every 3 months and 300 schools were evaluated at the end of the first year, even though, as I said before, the sample was not representative and the evaluation was based on interviews to parents, teachers and other members of the school community. More schools were benefited in subsequent years, but according to members of the Division of Educational Statistics, with whom I held a meeting, those new schools were not selected using a clear criteria, and therefore I do not include them in my analysis.

3 Data

The data I use in this paper come from the 2014 database of the Program for Schools of Excellence (PEE in Spanish), published by the Ministry of Education

(SEP); the 2015, 2016, 2017 and 2018 national results of PLANEA, and the 2010, 2011, 2012 and 2013 national results of ENLACE, the standardized test for primary and secondary schools that was used before PLANEA was implemented, published by the same Ministry. In 2014, as part of the transition to the new rules established by the Reform, no standardized test was implemented. I also use information from the Annual School Surveys, conducted by the Ministry for the same years for every public school in the country.

The PEE database contains information on the 32,615 public kindergarden, primary and secondary schools that were assessed to determine their eligibility for the program. They were selected using the information on schools located in marginalized areas from the 2013 Census of Schools, Teachers and Students from basic education (CEMABE), conducted by the National Statistics Institute (INEGI). It contains the ICE index calculated for every school, the infrastructural poverty classification according to the index, whether the school actually benefited from the program, the amount of money received for each component of the program and the main things done with that aid. It also contains information about the physical conditions of the school (access to water, bathrooms, type of floor, availability of chairs and tables for students and teachers, availability of blackboards, and whether it was built to be a school or not).

The databases from PLANEA contain information on the amount of students that the Ministry of Education planned to evaluate at each school, the number of students that were actually evaluated, and the percentage of students at each of the four levels of performance (where level 4 is the highest possible) for the 2 subjects evaluated: language and mathematics. The same is true for databases from ENLACE. I only consider students enrolled in the last grade of primary and secondary school (sixth and ninth grade, respectively), since after the Educational

Reform and the implementation of PLANEA those are the only ones subject to evaluation. Both tests evaluated all public and private schools recognized by the Ministry of Education. From 2010 to 2016 (except for 2014) both primary and secondary schools were evaluated, but the rules changed. Before 2014 all students were evaluated, however, for 2015 and 2016 the following rule was applied: where the number of students in sixth grade was less or equal to 35 all students were evaluated. If the number was greater than 35 but less or equal to 69, 35 students were randomly selected. If the number was equal to 70 two groups of 35 were formed and all students were evaluated. Finally, if the number was greater than 70 two groups of 35 were randomly selected and all students in those groups were evaluated. The same is true for ninth grade. In 2017 only secondary schools were evaluated, and every student enrolled in ninth grade was subject to evaluation. In 2018 only primary schools were evaluated, and again all students enrolled in the last grade were subject to evaluation.

Summary statistics for years 2010 through 2018 are presented in Table 1, 2 and 3. Table 1 includes all schools, whereas Table 2 and 3 include only primary and secondary schools, respectively. The PEE database includes information for the 32,615 public kindergarden, primary and secondary schools located in marginalized areas that were analyzed by the Ministry of Education in order to decide whether they were eligible for aid or not. Using the school code assigned by the government, and the school turn, I match schools in the program database to their correspondent results in PLANEA (or ENLACE) and the total amount of students registered at grades 6 and 9 (or third year of secondary school), for every year. I only take into account grades 6 and 9 because those are the ones evaluated by PLANEA. Since kindergardens are not evaluated, I do not take them into account. Total schools in PEE represents the total number of schools in the PEE database for which I want the results of PLANEA (or ENLACE) every year. To obtain it, I get rid of duplicates, of schools that had no students in grade 6 or 9 for that year and of those schools that appear neither in the census nor in PLANEA results for each year, since that means that those schools either closed or had no students in grade 6 or 9 (depending on whether it was a primary or secondary school).

Table 1: Percentage of schools matched (total)

	2010	2011	2012	2013	2015	2016	2017	2018
Total of schools in PEE	18,734	19,030	19,695	19,524	16,269	17,830	4,976	13,090
Schools that matched	16,880	16,880	17,635	16,990	9,035	14,618	4,319	10,159
Percentage lost in matching (schools)	9.8	11.2	10.4	12.9	44.6	18.0	13.2	22.39
Percentage lost in matching (students)	8.0	8.6	8.7	15.7	58.4	17.4	12.0	19.6

Notes: For every year this table summarizes the total number of schools that would ideally have test score results, the number of schools for which I have test score results, the percentage of schools lost and the percentage of students lost.

Table 2: Percentage of schools matched (Primary schools)

	2010	2011	2012	2013	2015	2016	2017	2018
Total of schools in PEE	14,346	14,534	14,802	14,195	11,258	12,821	-	13,090
Schools that matched	12,824	12,824	13,310	12,616	9,033	10,412	-	10,159
Percentage lost in matching (schools)	10.6	11.7	10.07	11.1	19.7	18.7	-	22.3
Percentage lost in matching (students)	10.6	11.5	10.9	17.4	15.5	20.6	-	19.6

Notes: For every year this table summarizes the number of primary schools that would ideally have test score results, the number of primary schools for which I have test score results, the percentage of primary schools lost and the percentage of students lost.

Table 3: Percentage of schools matched (Secondary schools)

	2010	2011	2012	2013	2015	2016	2017	2018
Total of schools in PEE	4,388	4,496	4,893	5,329	5,011	5,009	4,976	-
Schools that matched	4,056	4,056	4,325	4,374	2	4,206	4,319	-
Percentage lost in matching (schools)	7.5	9.7	11.6	17.9	99.9	16.0	13.2	-
Percentage lost in matching (students)	5.6	5.9	6.6	14.3	99.9	14.7	12.0	-

Notes: For every year this table summarizes the number of secondary schools that would ideally have test score results, the number of secondary schools for which I have test score results, the percentage of secondary schools lost and the percentage of students lost.

In matching schools to their test scores, more students are lost in primary schools than in secondary ones. As for the question of why are their PLANEA results missing, the reasons vary. For example, for 2016 indigenous schools make more than 30 percent of those that had no match, while another 38 percent is made up of schools that either closed or were distance-learning secondary schools (telesecundarias). As explained to me by staff at the department of Planning and Educational Statistics at the Ministry of Education, indigenous schools are hard to evaluate because many of them are not really schools: some are classes given at a room borrowed from the offices of the local government or some member of the community, and in some cases courses are given outdoors, due to lack to resources. On the other hand, distance-learning schools are harder to evaluate because there are no teachers, and they are more prone to decide not to participate in evaluations in response to unpopular government policies. As for the remaining third of schools, I consulted teh staff if the Ministry, who, after analyzing them with her team, concluded that there is no clear reason for their results not to exist, but it is probably due to unwillingness to participate in the test from those in charge of the schools.

Except for 2013 and 2015, the percentage of students that we lose after matching PEE schools with their grades in ENLACE or PLANEA is slightly lower than the same percentage in terms of schools, which tells us that these schools are not very large in general. The difference for 2013 and 2015, according to staff of the Ministry, can be explained by the low popularity of the Educational Reform implemented by the Mexican government in 2013. From Table 3 we have that most of the schools that decided not to participate are secondary schools. The Reform required significant changes in the Ministry, that included the design of a new standardized test. Therefore, schools were not evaluated in 2014, and when the new evaluation methods were used for the first time in 2015, they encountered significant resistance throughout marginalized areas of the country. The substantial decrease in the total number of schools in 2017 in Table 1 is due to the fact that only secondary schools were evaluated in that year, while only primary schools were evaluated in 2018.

Table 2 presents summary statistics for some baseline characteristics of the schools in the PEE database. It seems like having a bathroom is more common among secondary schools. The percentage of schools that do not have a blackboard stays roughly constant for all years at about one fourth, as well as the percentage of schools that have access to water (via tanker trucks or pipeline), at almost ninety percent.

Table 4: School characteristics

	2015	2016	2017	2018
Blackboard (%)	75.78	75.18	75.34	74.71
Bathroom (%)	58.31	55.99	61.11	52.84
No water (%)	21.05	13.28	12.29	13.48
Building adapted or no building (%)	43.27	42.8	46.1	41.77
Dirt floor (%)	6.82	8.77	8.46	8.61
Not every student has a chair (%)	43.13	41.31	45.31	41.06
Not every student has a table (%)	42.44	40.07	46.2	39.48
Not every teacher has a chair (%)	48.54	47.05	47.76	47.41
Not every teacher has a table (%)	44.54	43.18	43.75	43.5

Notes: This table summarizes observable characteristics of the schools that were considered to be part of the PEE program and have PLANEA results The total corresponds to schools that matched in Table 1.

4 Methodology

The Ministry of Education, based on the data of the 2013 Schools Census (CEMABE) elaborated by INEGI in 2013, created an index of infrastructural poverty, called ICE, that the government claimed to have used to select the schools that were going to be benefited by the PEE. The index supposedly divides schools into 5 categories of infrastructural poverty: very low, low, medium, high and very high. Only those schools classified with high or very high infrastructural poverty were eligible to participate in the program. The index takes into account 5 variables: type of building, building material, water availability, bathroom availability and basic equipment at the classroom.

After meeting with government officials from the Division of Educational Statistics, of the Ministry of Education, I was informed that the official cutoff for the program had been 0.24. This means that only those schools with an index equal or greater to 0.24 were eligible to receive aid. Since not all eligible schools actually benefited from the PEE, due to federal laws schools had to comply with, I cannot implement a sharp Regression Discontinuity design, but there is still a discontinuity, not a sharp but a fuzzy one, given by the probability of being treated. Schools below the threshold of 0.24 have zero probability of being treated, while schools above the threshold (or exactly at 0.24) have a strictly positive probability of being part of the program. This discontinuity in probability motivates the idea of using a Fuzzy Regression Discontinuity design to evaluate the program. To see whether the discontinuity exists I look at the relationship between the index and the probability of being treated.

Figure 1 shows the probability of receiving aid against the infrastructural poverty index (ICE), establishing 0.24 as the cutoff point, adjusted with a polynomial of order 2. Contrary to what we would expect, as the ICE index increases the probability of being treated decreases, up to around 0.6, and it increases beyond that point. This is in sharp contrast with the official eligibility rules of the PEE, and therefore makes the regression discontinuity design useless, since it relied on the assumption that the probability of being treated was not equal at both sides of the cutoff. In particular, the design relied on the idea that the probability of receiving aid increased as the ICE index increased.

Thus, I use a difference-in-differences design to evaluate the impact of the program. Following Angrist and Pischke (2009), I start by defining the outcomes of interest:

$$Y_{1it} = \text{percentage of students at the highest level of attainment at school } i \text{ and period } t \text{ if the school received aid from the PEE program}$$

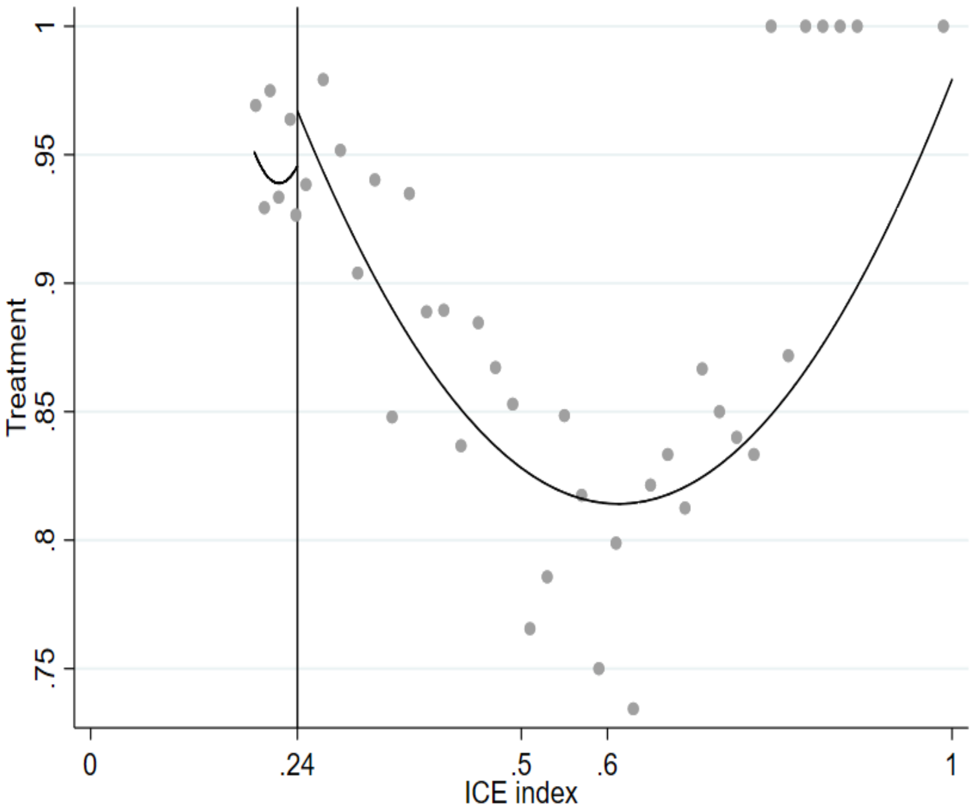
$$Y_{0it} = \text{percentage of students at the highest level of attainment at school } i \text{ and period } t \text{ if the school did not receive aid from the PEE program}$$

We should remember that this is the ideal case. In practice, we only observe one of the two outcomes for every school i. An important assumption is that, in the absence of a change in the school budget for infrastructure, the percentage of students in the highest level of attainment at school i is determined by the sum of a time-invariant school effect, θ_i, and a year effect that is common across schools, λ_t. Formally,

$$E[Y_{0it}|i,t] = \theta_i + \lambda_t \tag{1}$$

Now, let D_{it} be a dummy for schools receiving aid, where schools are indexed

Figure 1: The relationship between ICE and the probability of being treated



Notes: This figure shows the relationship between the index of infrastructural poverty for schools (ICE) and the probability of treatment. The index is continuous and takes values from 0 to 1, where 0 is no poverty and 1 means complete poverty.

by i and observed in year t. If we assume that $E[Y_{1it} - Y_{0it}|i, t]$ is constant, and we call that constant β, we can write the outcome as

$$Y_{it} = \theta_i + \lambda_t + \beta D_{it} + \epsilon_{it} \tag{2}$$

where $E[\epsilon_{it}|i, t] = 0$. Therefore, we have

$$E[Y_{it}|i = i_{control}, t = 2015] - E[Y_{it}|i = i_{control}, t = 2013] = \lambda_{2015} - \lambda_{2013}$$

and

$$E[Y_{it}|i = i_{treated}, t = 2015] - E[Y_{it}|i = i_{treated}, t = 2013] = \lambda_{2015} - \lambda_{2013} + \beta$$

where $i_{control}$ means that the school did not receive money from PEE in 2014, and $i_{treated}$ means that it received aid in that year. Thus, we can write the difference-in-differences coefficient as

$$\begin{aligned} &E[Y_{it}|i = i_{treated}, t = 2015] - E[Y_{it}|i = i_{treated}, t = 2013] \\ &\qquad - (E[Y_{it}|i = i_{control}, t = 2015] - E[Y_{it}|i = i_{control}, t = 2013]) = \beta \end{aligned}$$

In order to use a regression to estimate (2), and following the two period framework before, we simply let B_i be a dummy for whether school i was part of the PEE program and d_t a time-dummy for observations after the program was implemented, id est, 2015. In turn, we have

$$Y_{it} = \alpha + \theta B_i + \lambda d_t + \beta(B_i \cdot d_t) + X_i^\top \delta + \epsilon_{it} \tag{3}$$

where $(B_i \cdot d_t) = D_{it}$ and X_i is a vector of school controls that includes whether all classrooms have a blackboard, if there are are restrooms at the school, whether it has access to water, whether the school was built for educational purposes, whether they have dirt floor, whether all students have a table and a chair, and whether all classrooms have a table and a chair for the teacher. The vector of schools controls does not vary across time because information for the variables is given at one point in time, namely, 2013, and the variable that indicates whether the school was built for educational purposes does not vary. We now have

$$\alpha = E[Y_{it}|i = i_{control}, t = 2013] = \theta_{i_{control}} + \lambda_{2013} + X_{i_{control}}$$

$$\theta = E[Y_{it}|i = i_{treated}, t = 2013] - E[Y_{it}|i = i_{control}, t = 2013]$$

$$= \theta_{i_{treated}} - \theta_{i_{control}}$$

$$\lambda = E[Y_{it}|i = i_{control}, t = 2015] - E[Y_{it}|i = i_{control}, t = 2013]$$

$$=\lambda_{2015} - \lambda_{2013}$$

$$\beta = \{E[Y_{it}|i = i_{treated}, t = 2015] - E[Y_{it}|i = i_{treated}, t = 2013]\}$$

$$-\{\mathrm{E}[\mathrm{Y}_{it}|i = i_{control}, t = 2015] - E[Y_{it}|i = i_{control}, t = 2013]\}$$

Since we have data from 2010 to 2018 (with the exception of 2014) equation (2) can be easily generalized into

$$Y_{it} = \theta_i + \lambda_t + \sum_{l=-4}^{4} \beta_l D_{lit} + \epsilon_{it} \quad (4)$$

where D_{lit} are dummy variables equal to one if school i is l years after-treatment in year t, and the β_l are the coefficients of interest.

We can allow for pre and post-treatment tendencies to vary according to the vector of control variables, which results in

$$Y_{it} = \theta_i + \lambda_t + \sum_{l=-4}^{4} \beta_l D_{lit} + w_t^{\top} + \epsilon_{it} \quad (5)$$

where w_t is the vector of dummy variables for every year interacted with each control variable. Finally, following Angrist and Pischke (2009), and Zimmerman and Neilson (2014), we can use this equation to test whether the control and treatment group follow parallel trends before program implementation, which encourages one to think that the identification strategy is correct, given that we cannot test the post-treatment parallel trend in the absence of treatment assumption. So re-writing equation (5) we get

$$Y_{it} = \theta_i + \lambda_t + \sum_{\tau=1}^{4} \beta_{+\tau} D_{i,t+\tau} + \sum_{\tau=1}^{4} \beta_{-\tau} D_{i,t-\tau} + w_t^{\top} + \epsilon_{it} \tag{6}$$

where we expect $\beta_{-\tau}$ not to matter if our identification strategy is correct.

5 Results

Table 5 shows results using equation (4) and 2013 as the baseline year. Results are shown for the 2 highest levels of attainment (out of 4) in each subject, and standard errors are reported in parenthesis. I included as controls all other public schools that had results for ENLACE or PLANEA (depending on the year) and did not receive any aid from the government. For those schools, I obtain the control variables from CEMABE, the 2013 census that the Ministry of education used to construct the ICE index. It is clear that the difference-in-differences identification strategy is invalid here because for every regression the treatment and control groups do not follow parallel trends in at least 2 periods before PEE was implemented. The statistically significant coefficients (at 5 percent level) of periods prior to the program show that schools in the treatment group performed significantly worse in standardized tests than their counterparts. For instance, a school that was to receive aid from PEE had, on average, 5 percentage points less students in the highest level of attainment of mathematics in 2011 and more than 12 percentage points less students in the second highest level in the same year. The results in language are similar.

The after-treatment coefficients negative, but even when they are statistically significant -like in 2017 and 2018 for level 4 of mathematics- we are unable to draw any inference from them because the critical assumption of parallel trends is not satisfied.

Since the identification strategy did not work by only allowing for fixed effects at the school level, I turn to a specification where I allow for different trends in the observable characteristics of schools (access to water, availability of blackboards, tables and chairs for both students and teachers, etcetera) across years, before and after the treatment. Table 4 shows the coefficients of interest for equation (6), with 2013 as our baseline year and adding public schools that did not receive benefits of the program to the control group.

For level 4 of attainment in both subjects none of the coefficients prior to the implementation of the program is statistically significant at the 10 per cent level, which means that we cannot reject the hypothesis of parallel trends for the treatment and control groups. This allows me to use the coefficients after 2014 to estimate the effect of the program. If both groups were following parallel trends from 2010 up to 2014, an no other major event occurred in 2014 apart from the PEE, then a significant change after the implementation of the program must be due to it. In mathematics, the coefficients after the program are negative but none is significant, which at most enables us to say that there is no evidence of the program having an impact in that level.

In language, the coefficient for 2017 is negative and significant at the 1 percent level. Given that in 2017, due to new rules established by the government, only secondary schools were evaluated, reducing the number of schools to be matched from the PEE database to less than a fourth from what we have in 2015 and 2016 (see Table 1), the precision of our estimate increases. A negative effect is plausible. Improving the physical conditions of a school could mean, for example, construction workers present at the school for long periods of time in order to build restrooms, cement floors, new classrooms, repair tables and chairs. For instance, students from one grade might be forced to move and share one classroom with another group while theirs is being repaired, apart from being exposed to constant noise from the construction while trying to concentrate. This could seriously affect the learning outcomes of the teenagers that were in first grade of secondary school in 2014, and reached the evaluation year by 2017.

For level 3, both in mathematics and language it is still not clear that the difference-in-differences strategy works, since two of the diff-in-diff coefficients prior to the PEE are significant at the 5 percent level. Thus, as with equation (4), the difference-in-differences coefficients after the reform are useless to make any causal inference regarding the impacts of the PEE.

Table 5: Effect of improving infrastructure on test scores (without varying tendencies in school characteristics)

	Mathematics		Language	
	Students in level 4 of attainment	Students in level 3 of attainment	Students in level 4 of attainment	Students in level 3 of attainment
β_{-4}	-2.572	-10.992	-0.792	-16.131
	(1.34)	(1.70)**	(0.84)	(1.78)**
β_{-3}	-4.068	-13.548	-2.436	-17.646
	(1.32)**	(1.67)**	(0.82)**	(1.75)**
β_{-2}	-5.182	-13.425	-3.315	-16.694
	(1.39)**	(1.76)*	(0.87)**	(1.85)**
β_{+1}	-4.474	-4.280	-2.264	-11.634
	(1.78)*	(2.25)	(1.12)*	(2.37)**
β_{+2}	-2.258	-5.834	0.338	-8.126
	(1.31)	(1.66)**	(0.82)	(1.74)**
β_{+3}	-6.492	-3.203	-7.3	-10.621
	(2.06)**	(2.61)	(1.29)**	(2.74)**
β_{+4}	-5.789	-3.503	-1.806	-2.745
	(1.50)**	(1.90)	(0.94)	(1.99)
Constant	6.374	14.018	3.071	16.378
	(0.01)**	(0.02)**	(0.01)**	(0.02)**
Controls	Yes	Yes	Yes	Yes
N	462,547	462,547	462,524	462,524
R^2	0.09	.14	0.08	.18

Notes: Estimated coefficients for the impact of receiving aid from the PEE program on the percentage of students in the 2 highest levels of attainment at the school level for the 2 subjects evaluated, not allowing for different trends across years in control variables.$*$ denotes significance at the 95% level, $**$ denotes significance at the 99% level.

Table 6: Effect of improving infrastructure on test scores (allowing for varying tendencies in school characteristics)

	Mathematics		Language	
	Students in level 4 of attainment	Students in level 3 of attainment	Students in level 4 of attainment	Students in level 3 of attainment
β_{-4}	-0.826	-2.865	0.129	-3.458
	(2.11)	(2.67)	(1.32)	(2.81)
β_{-3}	-0.920	-5.862	-0.436	-6.282
	(2.05)	(2.59)*	(1.28)	(2.72)*
β_{-2}	0.924	-6.703	-0.679	-6.366
	(2.23)	(2.83)*	(1.40)	(2.96)*
β_{+1}	0.770	-18.861	-2.509	-7.297
	(4.89)	(6.19)**	(3.06)	(6.49)
β_{+2}	-2.391	-10.163	1.206	-4.840
	(2.11)	(2.67)**	(1.32)	(2.80)
β_{+3}	-3.363	-0.917	-4.048	-4.052
	(2.39)	(3.03)	(1.50)**	(3.18)
β_{+4}	-2.153	0.855	0.230	4.997
	(3.63)	(4.59)	(2.27)	(4.81)
Constant	4.754	11.563	1.656	10.326
	(0.51)**	(0.655)**	(0.32)**	(0.68)**
Controls	Yes	Yes	Yes	Yes
N	462,547	462,547	462,524	462,524
R^2	0.10	.14	0.09	.18

Notes: Estimated coefficients for the impact of receiving aid from the PEE program on the percentage of students in the 2 highest levels of attainment at the school level for the 2 subjects evaluated, allowing for different trends across years in control variables.$*$ denotes significance at the 95% level, $**$ denotes significance at the 99% level.

6 Conclusion

In 2014 the Mexican government implemented the first large scale program that assigned money directly to schools. It was part of the Educational Reform of 2013, intended to improve the quality of education throughout the country. The program, called PEE in Spanish, that benefited 20,000 kindergardens, primary and secondary schools, was aimed at improving the physical conditions of the schools and their management autonomy (the last one in a lower proportion). It was mandatory for schools to report on their use of the resources every 3 months and 300 schools that participated in the program were evaluated at the end of the year with apparently positive results, but even the authors of the evaluation stated that the sample was not representative.

This paper aims to bridge the gap by evaluating the effect of the program based on its impact on student test scores in standardized tests. The official lineups of the program established that the Ministry of Education was in charge of constructing an index of infrastructural poverty (ICE in Spanish) according to the school census from 2013, for those schools located in the most marginalized areas. The index would then serve to classify schools into one of five categories of poverty: very low, low, medium, high or very high. Only those schools ranked in the two highest levels of poverty were eligible to receive aid. The index cutoff was not made public, but after meeting with members of the Ministry I was told that the cutoff was 0.24, which initially lead me to think about the implementation of a regression discontinuity design. However, after analyzing the data according to such a design, I discovered that the probability of receiving aid for a school decreased as the index also increased, and that happened well beyond the official cutoff of 0.24. This invalidated the use an RD design, because a critical assumption was that there is an upward jump in the probability of being treated right after the cutoff.

Since the RD design was not useful in this case, I redesign the identification strategy to use a difference-in-differences methodology. Following Angrist and Pischke (2009), and Zimmerman and Neilson (2014), first I set up an equation including fixed effects at the school level, year effects and difference-in-differences coefficients for four periods before and after program implementation, using 2013 as the baseline year. I do so for the two highest levels of attainment of the two subjects evaluated on the tests. After finding that the critical assumption of parallel trends between the treatment and control groups was not satisfied, I turn to an equation allowing for different trends on observable school characteristics across time. I am then able to implement the difference-in-differences strategy at the highest level of attainment, both for mathematics and language. I find all but one of the coefficients after 2014 to be statistically non-significant, and the significant coefficient to be negative, providing evidence of a negative effect of the program on test scores. At level 3, the hypothesis of parallel trends is still rejected at the 10 percent level, invalidating the design.

The program had a budget of 7,600 million pesos, it was the main program of the Educational Reform and continued until the school year 2017-2018, right before a new government took power. However, measured by its impact on learning through student test scores in standardized tests 4 years after its implementation, the PEE program effect was, at most, insignificant. As for now, it seems like improving school infrastructure did not succeed in raising learning outcomes.

7 Bibliography

Glewwe, P., Hanushek, E., Humpage, S. and Ravina, R. (2011). School resources and educational outcomes in developing countries: a review of the literature from 1990 to 2010. *NBER Working Paper.*

McEwan, P. (2015). Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, 85, 353-394.

Jackson, C. K., Johnson, R. and Persico, C. (2015). The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms. *The Quarterly Journal of Economics*, 131, 157-218.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 91, 795-813.

Krishnaratne, S., Howard, W. and Carpenter, E. (2013). Quality education for all children? What works in education in developing countries. *International Initiative for Impact Evaluation (3ie) Working Paper.*

Newman, J., Pradhan, M. Rawlings, L. B., Ridder, G., Coa, R. and Evia, J.L. (2002). An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund. *The World Bank Economic*

Review, 16, 241-274.

Skoufias, E. and Shapiro, J. (2006). Evaluating the Impact of Mexico's Quality Schools Program: The Pitfalls of Using Nonexperimental Data. *World Bank Policy Research Working Paper.*

Valora Consultoría, S.C. (2015). Avances del Programa de Escuelas de Excelencia para Abatir el Rezago Educativo. *Informe financiado por el Banco Interamericano de Desarrollo.*

Hanushek, E. and Woessmann, L. (2011). The Economics of International Differences in Educational Achievement. *Handbook of the Economics of Education*, 3, 89-200.

Zimmerman, S. and Neilson, C. (2014). The effect of school construction on test scores, school enrollment, and home prices. *Journal of Public Economics*, 120, 18-31.

Angrist, J. and Pischke, J. (2009). Mostly Harmless Econometrics: An Empiricists Companion. *Princeton University Press.*