EVALUATING THE EFFECTS OF SCHOOL AND TEACHER INCENTIVES: QUASI-EXPERIMENTAL EVIDENCE FROM MEXICO

Edgar Franco Vivanco

May 2013

Abstract:

Merit-based incentives are a topic of growing interest in labor economics due to their potential to increase performance for private and public employees. Following this argument, such pay schemes have been applied in numerous countries to provide incentives to teachers and schools based on their students' achievement scores and other performance metrics. However, because of the multitask, multiprincipal and multiperiod nature of education, they present several caveats. Observational and experimental research provide ambiguous conclusions about their impact. This paper contributes to this literature by evaluating the effects of the Mexican PECD, a program that since 2010 has provided salary bonuses to teachers in primary and secondary public schools based on their national standardized tests scores. Nearly 30,000 schools and 300,000 teachers have benefited from this program in its first implementation. Given its characteristics, the PECD provides an ideal ground for a Regression Discontinuity Design (RDD) and for a Multiple Rating-Score Discontinuity (MRSD). Combining these quasi-experimental techniques with a Hierarchical Linear Model (HLM), I show that the effect of receiving an award on student performance is null. If any, this program seems to harm student performance; this negative effect is more evident for indigenous schools. Some possible explanations of these results are discussed here.

1 Introduction

The aim of incentives is to change existing behavioral patterns and motivate new ones. The rationale to apply incentives in the working environment is based on the conception that mechanisms can be designed to compensate workers according to some observable outputs and, by doing so, generate higher effort in the determined tasks (Holmstrom and Milgrom 1991). This same assumption is also behind the proposed incentives for public sector workers (Dixit 2002) and teachers—as the main labor input in the school production function (Ballou 2001; Lavy 2007). However, implementation of incentives for public schools is problematic because education is complex and the public education system is a "multitask, multiprincipal, multiperiod, near-monoploy organization with vague and poorly observable goals" (Dixit 2002, 719). It is "multitask" because it is intended to achieve several objectives, including imparting cognitive and non-cognitive skills, fostering the growth of the children, and preparing students to enter the work force, among others. It is "multi-principal" because there are several stakeholders acting in overlapping relationships, including teachers, parents, students, taxpayers, policymakers, potential employers of the graduates, teachers' unions and the society as a whole. It is "multi-period" because teachers learn to teach over time and often after many years of practice. It is quasi-monopolistic because, although there are private options, most of the population depends on the public sector for the provision of education. And their goals are vague because agents (teachers) enter the profession often highly motivated by non-pecuniary factors. In democratic countries additional problems of agency might be involved between voters and public officials designing those schemes (Dixit 2002). The presence of any one of those elements causes difficulties in the implementation of an incentive system. Yet the presence of all of them, simultaneously, truly complicates the design of a proper and universal scheme of incentives in the educational sector.

Despite the difficulty of implementation of these schemes, they remain an increasingly popular policy in several countries. No surprisingly, there is ambiguous evidence when measuring the impact of such programs on different outcomes, particularly on student achievement. Inconclusive results from incentive programs for teachers may be due in part to the lack of information regarding how to design an effective incentive program in this domain. At the same time, incentives assume that the problem with low student performance is due to lack of teacher effort and motivation, even though there are reasons to believe that this problem is more related to teacher competence (Springer et al. 2010). There is a growing body of literature that studies the effect of many different incentive programs across the world. In the United States, incentive programs have been implemented in North Carolina (Vigdor 2009), Dallas (Ladd 1999), Kentucky (Koretz and Barron 1998; Koretz 2002), Florida (Kenny and Figlio 2006), Little Rock (Winters et al. 2008) and Denver (Gonring 2007). Internationally, there is evidence from Israel (Lavy 2002, 2005), England (Atkinson et al. 2009), Portugal (Martins 2009) and Chile (Mizala and Romanguera 2004).

Most recently, some randomized control trials have been conducted to better understand the causal effects of teacher and school incentives on student performance. Again, results are inconclusive. Evidence from India shows a positive impact of incentives (Duflo and Hanna 2005; Muralidharan and Sundararaman 2010) whereas evidence from Kenya (Glewwe et al. 2010) and United States (Springer et al. 2010; Goodman and Turner 2010; Fryer 2010, 2011) shows no effect. In any case, it is clear that more evidence is needed to provide a more compelling verdict about these schemes.

This paper studies the case of the Mexico's *Programa de Estímulos para la Calidad Docente* (PECD),[1] established in 2010 with a budget of 899 million pesos (74 million USD) with the objective of providing economic incentives to teachers and schools based on student performance. This program uses scores from the national standardized tests *Evaluación Nacional de Logro Académico en Centros Escolares* (ENLACE)[2] to rank schools according to their achievement and improvement. After grouping schools in homogeneous clusters within each of the 32 states, it provides incentives to schools in the top 15 percent of performance or improvement. In total, there are more than 500 different clusters for within comparison. Each of the more than 120,000 primary and secondary public schools that have administered ENLACE for a minimum of three years is eligible for this program. In 2010, about 30,000 schools received a reward that was then distributed evenly to nearly 250,000 teachers. These rewards account for about four percent of their yearly salary. The program has continued to date with some modifications and an increased budget; the present research is the first impact evaluation of this program. More details about this program are provided in section 3.

The PECD provides an ideal ground for a Regression Discontinuity Design (RDD) because it establishes exogenous cutoffs that allow for a comparison between schools near the boundary of the receipt of treatment. In order to control for fixed effects of the different clusters while allowing a variation of the effect of receiving the incentive, I specified a

[1] Programm for Teaching Quality Incentives
[2] National Evaluation of Academic Achievement in Scholar Centers

Hierarchical Linear Model (HLM) that allows a random slope for the treatment effect. In order to further analyze heterogeneous effects, I restrict the population to several subsamples of schools. Additionally, to test the effect of receiving the two different incentives distributed by the PECD I used a Multiple Rating-Score Regression Discontinuity (MRSRD) as proposed by Robinson and Reardon (2010) and Imbens and Zajoc (2011). The results show no significant effect of receiving any of PECD incentives on the ENLACE test scores for 2011, nor for the school value added -measured as the change in the standardizes score between 2010 and 2011. There are also no effects for receiving both incentives. Moreover, the PECD seems to have a harmful effect, which is more persistent for indigenous schools.

This paper contributes to the literature by providing evidence of the effect of a national level incentive program in a middle income country. Because programs at these levels are relatively new, there is little information about their impact at large scale (Martins 2009). The main advantage of this research is that it uses a quasi-experimental approach, a large N and a clustering criteria at the same time, thus tackling issues of causal inference and external validity simultaneously. By doing so, this paper helps to enhance the understanding of teacher and school incentives based on performance. Here is important to clarify that because all schools entered to the program at the same time there is no counterfactual to test the effects of being offered the incentive, in this sense this paper estimates the effect of receiving an award rather than the effects of incentives.

The organization of the paper is as follows: In next section I provide a review of the literature on the effect of teacher and school incentives on student achievement. In section 3, I provide a description of the PECD and the particular characteristics of the first round of the program. In section 4, I describe the data analysis and the empirical strategy. Next I describe the results and discuss their relevance. Finally, section 7 provides conclusions.

2 Economic incentives in the literature

The literature on the role of teacher and school incentives on student performance is a growing area of interest in education and labor economics. Studies on the topic have been conducted in many countries using observational data and, most recently, experimental designs. There is also a nascent interest in the optimal design of such payment schemes (Neal 2011). In general, the evidence about the impact of teacher incentives on student achievement is ambiguous. A more general consensus is that the structure of these incentives is crucial to obtaining the expected results, and that incentives might be regarded more as a complement rather than a substitute to other policies oriented to increase student achievement and quality of instruction (Neal 2011).

In the United States, incentive programs have been implemented in North Carolina (Vigdor 2009), Dallas (Ladd 1999), Kentucky (Koretz 2002), Florida (Kenny and Figlio 2006), Little Rock (Winters et al. 2008) and Denver (Gonring 2007). There have also been programs implemented across states (Hudson 2010). For some researchers, these programs are positively regarded, since the little variation in teachers' salaries across the country might be driving many high-aptitude individuals out of the teaching profession (Hoxby and Leigh 2004). The evidence of the impact of these implementations is, overall, positive, but this should be taken cautiously given the existence of some unintended effects and the non-experimental assignment of incentives. For example, Vigdor (2009) shows that the North Carolina program has large gains in math and reading but given the NAEP trends the reading results are suspicious and might be representing artificial score inflation. This situation is somehow replicated in the Kentucky bonus program, where students score higher on the state assessment but have no improvement in the NAEP test (Koretz 2002). On the other hand, Ladd (1999) find a positive effect of the Dallas program but only for white and Hispanic students, thus increasing the learning gap for black students.

In other countries, Lavy (2002) studies a program implemented in Israeli high schools using an assignment error in the program as a natural experiment. Corroborating with alternative methods of matching and regression discontinuity, Lavy finds that monetary incentives to teachers have a positive effect on student achievement. On the contrary, a program implemented in Portugal (Martins 2009) and one implemented in England (Atkinson et al. 2009) show no evidence of positive effects on student achievement. Martins (2009) uses a matching and a difference-in-difference method to analyze the effect of a program implemented in Portugal public sector schools at national level from 2002 to 2008. The author finds a significant and negative effect of receiving the incentives on student achievement and finds a significant increase of score inflation. Atkinson et al. (2009) examine the Performance Related Pay England, using a quasi-experimental variation; the authors find a positive effect of the scheme on English and Science value added, but no effect for Math teachers. A common characteristic for Portugal and England programs is that they rely on subjective assessment by peers or officials which might contribute to the relative ineffectiveness of these programs (Neal 2011).

In Latin America, the case of Chile is particularly interesting for the present study. Since 1996 Chilean authorities have implemented the *Sistema Nacional de Evaluación de Desempeño de los Establecimientos Educativos* (SNED), a program with similar characteristics to the Mexican PECD. The SNED groups together schools of similar characteristics, then ranks them according to different criteria and provides an incentive to the best performing schools in each region until it reaches 25 percent of the region's enrolment. The incentives are distributed every other year and account for five to seven percent of teachers' annual salary. Mizala and Romanguera (2004) provide an assessment of this program and its impact on student achievement. They find that there are positive and significant effects of the SNED school effectiveness when three applications are considered in the same model. However, the methods used by these authors do not completely provide a valid counterfactual to assess its impact.

The increasing interest in merit-based programs has influenced experiments to randomly assign treatment and control groups to provide a more compelling causal inference on the impact of these programs. The available experimental evidence on this topic is still scarce and unclear but provides experiences from very different settings: Kenya (Glewwe et al. 2010), India (Duflo and Hanna 2005; Muralidharan and Sundararaman 2010), Nashville, Tennessee, (Springer et al. 2010) and New York City (Goodman and Turner 2010; Fryer 2011).

Glewwe et al. (2010) analyze a program that took place in 50 (out of 100) randomly selected Kenyan primary schools in 1998 and 1999. Incentives were distributed to schools with the largest scores or showing the highest improvement and then divided equally among teachers from grade 4th to 8th. There were also penalties associated for missing the exam. The authors find a positive effect in treatment schools on government exam participation and on second year scores by 0.14 standard deviations. These results, however, were not maintained once the program was completed and the scores of tests were not linked to incentives (applied by a NGO) did not increase. For Glewwe et al., these results can be interpreted as an evidence of coaching which is not related to long-term learning and human capital formation.

Duflo et al. (2005) combined a randomized experiment and a structural model to test the effect of monitoring and financial incentives on teacher absence and student performance in rural India. The authors find a positive effect of the incentives on the decision of teachers to attend the school that cannot be only related to monitoring; students in program schools benefited from about 30 percent more instruction time. The program also presented a statistically and economically significant impact on test scores of about 0.17 standard deviations higher for program schools. Additionally, the authors conclude that the program was cost-effective.

Muralidharan and Sundararaman (2010) present the results of a randomized evaluation of a program that provided bonus payments to teachers based on the improvement of their students' tests scores in the Indian state of Andhara Pradesh. In this program, 100 schools received group bonuses based on school performance and 100 received the bonuses based on individual performance. After two years, students in schools that received the incentive performed significantly better by 0.28 standard deviations in math and 0.16 standard deviations in language. Although group and individual incentive schools performed equally well during the first year of the program, individually incentivized schools performed better in the second year.

In United States, Springer et al. (2010) evaluated the Project on Incentives in Teaching (POINT), a three-year study conducted in the Metropolitan Nashville Schools System from the 2006-2007 through 2008-2009 school years. In this project, mathematics teachers voluntarily participated in a program that rewarded them with bonuses up to $15,000 if their students showed large gains in standardized tests. Almost 300 teachers (70 percent of all-middle school math teachers in the Nashville metropolitan area) participated of their own volition. The results of this study are that financial incentives do not contribute to increased student scores. The authors conclude that incentives do not affect student outcomes if they are implemented alone. Without professional development and a formal method for training on how to improve scores, incentives have no effect on increasing student scores.

Goodman and Turner (2010) and Fryer (2011) investigate the NYC Schoolwide Performance Bonus Program that randomly assigned low performing schools to treatment and control groups. Treatment schools provided bonuses up to $3,000 to teachers that increased student achievement. Goodman and Turner (2010) find a negative and small impact of the program on student achievement but also a small reduction in absences of teachers receiving the largest incentives. On the other hand, differentiating for Treatment on the Treated (TOT) and Intention to Treat (ITT) effects Fryer (2011) did not find any effect on student achievement, and also consider the possibility of a reduction on test scores for students attending three years in a treatment school of 0.045 standard deviations in math and 0.033 in reading. Fryer finds the impacts on other outcomes (attendance, behavioral incidence, etc.) to be negligible.

As presented here, the evidence either using observational or experimental data of the effect of financial incentives on learning is ambiguous and varies from country to country. The present paper contributes to the literature by providing evidence of the effects of a nationwide program in a quasi-experimental setting. Contrary to experimental designs, this research is not constrained by a small number of observations. About 120,000 schools were eligible for the reward and this was distributed to approximately 250,000 teachers in nearly 30,000 schools; this large sample allows for more accurate inferences with more external validity and the possibility to interact with different subsamples. As the experimental studies cited above were conducted in either low income countries or the United States, this research can also help to better understand the effect of these schemes in a middle income country.

3 The Programa de Estímulos a la Calidad Docente (PECD)

Mexico has a large and complex public educational system that includes more than 200,000 basic education schools, more than one million teachers, and about 25 million students (OECD 2011). It is divided by type of school according to community size, location and ethnic background (INEE 2010). Recently, Mexico has embarked on an effort of educational reforms to enhance educational quality and improve the level of instruction. Teacher evaluation and modification of their payment schemes has been an important element of such efforts. The last large reform in that area took place in 1993 through the *Carrrea Magisterial Programme*, which provides supplements to teacher salaries and possibility of promotion after meeting various criteria. Despite being highly promising, this program has shown small effects on student achievement (Santibañez et al. 2006)

The PECD was introduced in Mexico in 2010 as part of a set of policies contained in the *Alianza por la Calidad de la Educación*. The agreement was signed on May 15, 2008, by the Ministry of Education and the teachers union, with the objective of "transform[ing] the educational model through public policies that enhance quality and equity in education" (Amador Hernández 2009, 1). Specifically, the aim of the PECD is "to incentivize teaching practices that contribute to enhanced educational quality." (SEP 2010a).

The PECD provides financial incentives to primary and secondary public schools based on their ENLACE scores, which relied on four criteria designed by the National Institute of Educational Evaluation (SEP 2010a):

1) The primary and secondary schools with the average cumulative scores in Spanish and Mathematics in the top 15 percent of scores for the period 2006-2009 (High Achieving). [3]
2) The primary and secondary schools whose improvement in test scores falls in the top 15 percent, when improvement is measured by comparing scores from 2009 to the average of scores from all former years (Most Improved).
3) Teachers from targeted classes in primary schools (general, indigenous and communitary) and secondary schools (communitary and distance learning) whose average students' scores in Spanish and Mathematics for 2009 are at the top 15 percent on the test and their schools do not fall into categories 1 and/or 2.[4]
4) And teachers from targeted classes in secondary schools (general, technical and workers oriented) whose average students' scores in Spanish, Mathematics and the third subject evaluated in 2009 (Civic education) are in the top 15 percent on the test and their schools do not fall into categories 1 and/or 2.

For criteria 1 and 2, the money was distributed at the school level and then distributed evenly among teachers, principals and vice principals of each school while bonuses based on criteria 3 and 4 are given directly to the targeted students' teachers. During the first round of PECD in 2010, eighty-seven percent of the total amount of monies were distributed to the schools that fulfilled requisites 1 (High Achieving) and 2 (Most Improved) described above, and therefore this research is focused on these rewards at the school level.

To determine the ranking, the PECD relies on a grouping system within each state so that schools compete for rewards against comparable schools (SEP, 2010b). Four types of criteria are used in grouping schools: 1) Level of school (primary or secondary); 2) the type of school (general, indigenous, communitary, technical, or distance learning); 3) the municipality's socioeconomic status; [5] and 4) the location of the school (rural or urban). The aim of assigning PECD incentives by group of schools is to maintain a level of fairness by

[3] If there is not information for the four years, at least three are required.

[4] In Mexico, schools are classified based on their structure and administrative characteristics. Primary schools can be general (regular education), indigenous or communitary; secondary schools can be general, workers oriented, communitary, technical or distance learning. Communitary schools, henceforth CONAFE, are schools administered by a decentralized and autonomous public entity, the *Consejo Nacional de Fomento Educativo*. These schools are established in small communities that do not meet the minimal criteria established by the Ministry of Education in terms of number of students.

[5] As defined by the *Consejo Nacional de Población* (CONAPO). The PECD uses a two-tier classification system, where municipalities with Low and Very Low level of vulnerability are merged in one group and the rest (Medium, High and Very High) into a second one.

comparing schools with similar characteristics and reducing performance differences among schools associated with the disparity in the socioeconomic levels of students. After the schools were grouped together, the cutoffs for the top 15 percent of scores is set. Only the schools above this cutoff received rewards. The classification criteria applied for each of Mexico's 32 states are shown in Figure 1. Here we can observe that a state could potentially have up to 28 homogenous clusters.

[FIGURE 1]

For the first round of incentives in 2010, a total of 899 million pesos (74 million USD) were distributed among 259,014 teachers, principals and vice principals in 28,187 schools. The average amount distributed per school was 27,782 pesos (2,315 USD), or about 3,500 pesos per teacher (281 USD). The total amount of monetary rewards is a function of teachers per school and average ENLACE score of the clustering of schools (SEP 2010) and was designed so the amount per teacher could not be less than 2,000 pesos (166 USD) pesos and no more than 20,000 pesos (1,600 USD).

Table 1 shows the number of primary schools rewarded by the PECD in each state either for being High Achieving, Most Improved or both. The frequencies show that, on average, around 15 percent of the schools were rewarded by each of these mechanisms, as the design of the program indicates. With the exception of Michoacan and Oaxaca—where respectively only 4.9 and 2.6 percent of the schools were rewarded for being Most Improved and 4.5 and 2.4 percent for being High Achieving—the variations can be explained by adjustments within groups to exclude schools with suspicious results. [6] Another element that Table 1 shows is the number of schools that received both rewards; in total only about 3 percent of primary schools received an incentive for having a high score and for achieving large gains. Additionally, Table 1 includes a column showing the number of clusters defined for each state; this number ranges from 7 to 25 depending on the size and complexity of each state's school system and socioeconomic characteristics. That means that no state fills all 28 potential clusters shown in figure 1. Estado de México has the most complex educational

[6] It is very likely that this imbalance in Oaxaca is due to the high number of schools that refuse to be evaluated by ENLACE every year in that state. Schools from Oaxaca are part of the Section 22 of the Teacher's Union that has been boycotting the Alliance for Quality Education since the beginning of this program in 2008; a similar situation is also present in Michoacán (La Jornada 2008).

system according to this classification scheme with 25 different groups, while Distrito Federal (an urban homogenous area) has only 7.

[TABLE 1]

4 Data and Empirical Strategy

The data used in this paper comes from the first round of PECD assignation that took place in 2010. After the completition of this round, the Mexican Ministry of Education published the schools that received the awards and the total amount assigned to them. To date, in an exercise of transparency, nine states and all CONAFE schools have also made public the names of teachers and copies of the checks they received.[7] The data that was not available through the webpage, that is, the rating scores, the cutoffs for each group and the grouping of schools that did not receive the award, was solicited through standard mechanism of public information requests.[8]

The outcome data used in this research is the ENLACE score for 2011, one year after the distribution of rewards. ENLACE is the standardized national test applied in Mexico to all primary and secondary schools since 2006. This tests covers two subjects every year (Spanish and Mathematics) and an additional subject that changes year to year. Here I use the average score for Spanish and Mathematics at school level which was then standardized to obtain the z-score.[9]

A second outcome studied here is the change in the normalized scores for each school from 2010 to 2011, which can be characterized as a school value-added.[10] Additionally, in order to control for pre-treatment characteristics, I include a control for the size of the school, measured by the number of students and characteristics at school level. These characteristics were compiled from answers to questionaries from students, teachers, parents and principals applied at the time of the ENLACE tests. Here, I use the ones administered to principals because those are the only ones available for all schools while the others were only completed by a sample of schools. Since the first round of PECD incentives was distributed right after the

[7] http://estimulosalianza.sep.gob.mx/

[8] Through a petition to the *Instituto Federal de Acceso a la Información y Protección de Datos* (IFAI)

[9] Alternative models used Mathematics and Spanish score by separate with no meaniningful different results compared to the average score.

[10] Given the limitations that exist in Mexican data to track students through time, this research is restricted to a schoo-level analysis.

application of ENLACE 2010, it is possible to use the data from the questionnaires of this year as a control for pretreatment characteristics.

As those questionnaires are extensive and include several questions with high redundancy, in order to avoid overspecification of the model, I built nine indexes at school level using a principal component analysis and estimating a new variable based on varimax rotated factors. The indexes used and the variables included in each of them are shown in Table 2.

[TABLE 2]

4.1 Regression Discontinuity Design (RDD) in a Hierarchical Linear Model (HLM)

The advantage of using a RDD is that, in the absence of randomized control trials, it can produce plausible estimates of the treatment effect on an outcome for individuals near the cutoff score (Robinson and Reardon 2010). That is, individuals not receiving the treatment which are close to the cutoff provide a valid counterfactual to individuals receiving treatment. In this case, individuals are schools, the treatment effect is receipt one of the two incentives (High Achieving or Most Improved); and the outcomes are the standardized scores for 2011 and the school value-added from 2010 to 2011. Table 3 presents a balance check of pre-treatment conditions for schools near the cutoff. Note that given the nature of the PECD, the accurate way to test this balance would be for each homogeneous cluster or a group of them; for example, by comparing general primary schools in poor and rural areas only. As shown in table 3, schools near the cutoff not receiving the incentive can be used as a valid counterfactual for schools near the cutoff receiving the incentive.

[TABLE 3]

The application of RDD in education has a long history that can be tracked to the work of Thistlewaite and Campell (1960) on the effect of a National Merit Award and the study of Campbell and Stanley (1963) on teaching. Recently, there has been an increasing interest in this methodology to solve selection bias using observational data (Angrist and Lavy 1999; Black 1999; Kane 2003; Reardon et al. 2010). The present analysis builds in this literature and provides some elements of further discussion given two important particularities of the PECD: the existence of 544 unique clusters, each one with two cutoff scores determining two treatments.

The PECD provides an ideal setting for using a RDD to evaluate the effect of school incentives because the cutoffs established in 2010 were calculated with scores from 2006 to 2009, and there is no reason to believe that schools had any knowledge in previous years of how this program would be designed nor the position in the rating they would have. In this regard, the implementation of the cutoff and the incentive assignation can be considered exogenous, which is the main assumption for RDD. In order to verify this assumption, as suggested by Lee (2008) and Lee and Lemieux (2009) I tested the smoothness of the densities of the treatment-determining variables –rating variables– and found no evidence of a potential manipulation of the treatment assignment. Then, establishing a simple regression:

$$y = \alpha + f(X) + \gamma T + \varepsilon$$

$$T = 1[X > cutoff]$$

The potential relationships between the outcome (Y) and the treatment (T): $E[y|T = 1, X]$ and $E[y|T = 0, X]$ are both continuous in the rating variable (X). If the potential error is also continuous, $E[\varepsilon|T = 0, X]$, then the difference in conditional expectations is the treatment effect at the cutoff:

$$\lim_{X\downarrow cutoff}[y|X] - \lim_{X\uparrow cutoff}[x|X] = \gamma$$

Another assumption of the RDD is compliance of treatment. That is, that all individuals below the cutoff do not receive the treatment and that all individuals above the cutoff do. In the case of PECD none of the schools under the cutoff was assigned to receive an incentive, although some schools above the cutoff did not receive one.[11] though the reasons are not completely clear, it is possible to infer that these schools were catalogued with suspicious results by Mexican authorities because they present extremely high scores and increases compared to the other schools in the cluster. As these schools are outliers, they do not fall within the bandwidths specified in the models. However, in order to avoid a potential problem of non-compliance, they were excluded from the analysis.

[11] Since federal authorities transferred these resources to state authorities, it is only certain that schools in the nine states that reported the distribution of these resources received their incentives. For the purposes of this research, I assume that all schools assigned to an incentive received it.

In order to provide a common cutoff for all schools, a new rating variable was created using the difference of each school rating score and the cluster cutoff. Figures 2 to 5 show the relationship between the new rating variables and the average of the two outcomes studied here using a bin of 5 points.

[FIGURE 2]

[FIGURE 3]

[FIGURE 4]

[FIGURE 5]

A first visual inspection provides no evidence of an effect of the treatments at the discontinuity and indicates that the best fit of the functional form might be a linear model in the neighborhood of interest. My first approach to the functional form was, however, agnostic and allowed flexibility to the relationship between rating variables and outcomes. Thus the model is:

$$y_{i,c} = \alpha + f(X_{i,c} - X_c) + \gamma T_{i,c} + \mu S_{i,c} + e_{i,c}, \text{ where } \{X_{i,c} - X_c = \tilde{X}_{i,c}\} \in N \qquad [1]$$

Where for each school i in a cluster c the outcome score (y) is determined by a function of the rating score $(\tilde{X})$ in a given neighborhood N of real numbers close to the cutoff, and the a binary variable indicating if the school was assigned to treatment (T). That is, $T_{i,c} = 1[X_{i,c} \geq X_c]$, where X_c is the cluster cutoff. Then, by centering all cutoffs at zero, the new frontier represents the intercept point. The coefficient of interest is thus γ, which is the average treatment effect. Additionally, a vector of pretreatment characteristics of the school (S) can be included to increase the precision of the estimates, although generally this inclusion is not necessary (Imbens and Lemieux 2008). The neighborhood used in this paper goes from a bandwidth of 40 to 15 points on each side of the cutoff. This neighborhood accounts for between 30 percent and 10 percent of observations and represents less than a half of a standard deviation for both rating variables.

A particular element of interest in this research is the possibility of differentiated effects of the treatment in each cluster. For this reason, I provided a modified version of model 1 fitting the following two-level Hierarchical Linear Model (HLM):

$$y_{i,c} = \alpha_c + f(\tilde{X}_{i,c}) + \gamma_c T_{i,c} + \mu S_{i,c} + e_{i,c}, \text{ where } \{\tilde{X}_{i,c}\} \in N \quad [2]$$

And:

$$\alpha_c = \alpha_0 + u_c$$

$$\gamma_c = \gamma_0 + v_c$$

α_c refers to the intercept in cluster c

α_0 refers to the overall intercept, or the grand mean of the outcome variable across all clusters when all predictors are equal to zero

u_c refers to the random error component for the deviation of the intercept of a cluster from the overall intercept

γ_c refers to the slope in cluster c for the relationship between school level treatment and outcome score

γ_0 refers to the overall regression coefficient

v_c refers to the error component for the slope

Using a HLM approximation allows for random intercepts and random slopes for treatment effects at cluster level. The key element of HLM is that these random deviations are different than those associated with the overall error term (Raundenbush and Bryk 2002). These models provide the standard deviation of the estimate across different clusters. Again, the model is restricted to a neighborhood close to the cutoff. A further step in this analysis includes the identification of heterogeneous effects for different types of schools by restricting the model to different samples.

4.2 Response Surface Regression Discontinuity (RSRD)

A characteristic of the PECD is that it provides two types of incentives. In terms of a RDD this means that there are two different cut-offs. These can be observed in Figure 6, where quadrant *a* shows the schools receiving two incentives, quadrant *b* shows schools receiving incentive 2 (Most Improved), quadrant c shows schools not receiving any incentive and quadrant *d* shows schools receiving incentive 1 (High Achieving). The different boundaries define the possible discontinuities in this data.

[FIGURE 6]

Following Robinson and Reardon (2010), the effect of receiving multiple treatments can be tested using a Multiple Rating-Score Regression Discontinuity (MRSRD). In particular, to estimate the effects of each treatment relative to other treatments, is possible to use a RSRD. That is, assuming that treatment effects do not vary systematically within each rating score along both cut-offs, we can compute the following model:

$$y_{i,c} = \alpha_c + f(\tilde{X}^1{}_{i,c}) + g(\tilde{X}^2{}_{i,c}) + \gamma_c T^1{}_{i,c} + \lambda_c T^2{}_{i,c} + \mu S_{i,c} + e_{i,c}, \text{ where } \{\tilde{X}^1{}_{i,c}, \tilde{X}^2{}_{i,c}\} \in N$$
[3]

And:

$\alpha_c = \alpha_0 + u_c$
$\gamma_c = \gamma_0 + v_c$
$\lambda_c = \lambda_0 + k_c$

Here, $\tilde{X}^1$ and $\tilde{X}^2$ represent the pooled rating score of High Achieving and Most Improved respectively, for school i in cluster c. Now, coefficients of interest are γ_c and λ_c, which also have a random component at cluster level (v_c and k_c respectively). The advantage of this model is that it allows the use of all available data providing relatively precise estimates of treatments.

4.3 Frontier regression discontinuity (FRD)

According to Robinson and Reardon (2010), a disadvantage of the former method is the necessity of strong assumptions regarding the functional form. In order to avoid such complexities in the model it is possible to restrict the model to a subset of data on all but one of the rating scores. In this case, we can restrict the sample to all schools that received treatment 2 (Most Improved), and compare the schools near the cut-off for receipt of treatment 1 (High Achieving). In other words, we are comparing the schools that were close to receiving both incentives but only received one to those that received both. More formally:

$$y_{i,c} = \alpha_c + f(\tilde{X}^1{}_{i,c}) + g(\tilde{X}^2{}_{i,c}) + \theta_c T^{1\&2}{}_{i,c} + \mu S_{i,c} + e_{i,c}, \text{ where } \{\tilde{X}_{i,c}\} \in N \text{ and } [T^2{}_{i,c} = 1]$$
[4]

And:

$\alpha_c = \alpha_0 + u_c$
$\theta\gamma_c = \theta_0 + v_c$

In terms of the Figure 6 the estimate θ_c now represents the average effect of receiving two incentives versus receiving only incentive 2 for schools along the frontier $R_{a|b}$. The inclusion of the other rating scores and the covariates can be used to increase the precision of the model.

Equivalently, it is possible to fit the model for the subsample of schools receiving incentive 1 and estimate the effect of receiving two incentives versus only one incentive along frontier $R_{a|d}$ of figure 6. Formally:

$y_{i,c} = \alpha_c + f(\tilde{X}^1{}_{i,c}) + +g(\tilde{X}^2{}_{i,c}) + \theta_c T^{1\&2}{}_{i,c} + \mu S_{i,c} + e_{i,c}$, where $\{\tilde{X}_{i,c}\} \in N$ and $[T^1{}_{i,c} = 1]$ [5]

And:

$\alpha_c = \alpha_0 + u_c$

$\theta_c = \theta_0 + v_c$

Note that model 4 and 5 also allow for a random constant and slope of at cluster level allowing to determine the standard deviation of the coefficient along frontiers $R_{a|b}$ and $R_{a|d}$. It is also relevant to highlight that the estimates for each subpopulation are not necessarily comparable because effects are not necessarily homogeneous.

5 Results

All tables below show the results for the pooled cutoffs using a bandwidth from 40 to 15 and estimating two different outcomes. For the sake of simplicity, only the coefficients of interest determining the potential change in the outcome variable at the frontier are shown here along with their standard deviations from the random model. In all cases, the covariance specified for the mixed models was unstructured.

5.1 Effects of receiving the award

Table 4 provides estimates for model 2. The different rows show different specifications with and without controls. Although I estimate other models using quadratic terms and more flexible functional forms, the results are not presented here because they do not change meaningfully compared to the linear model. The results show no significant impact of receiving incentive 1 (High Achieving) on the standardized score 2011, nor on the school value-added. Although on average it is slightly negative, the standard deviation provided by the random model shows that in some cases this estimate could be positive for certain clusters.

[TABLE 4]

The effect of receiving the second incentive (Most Improved) over the 2011 score of the school is negative and significant at 95 percent for a bandwidth of 40 points and significant at 90 percent for all but one of the other bandwidths studied. The standard deviation of the coefficient provided by the random model is small enough to indicate the persistence of this negative effect across groups. However, this effect is relatively small (around 0.03 standard deviations) and it vanishes once pre-treatment controls are included. The effect is also not significant for the impact of this incentive on school value-added.

Figures 7 to 10 show the estimate of the effect of both treatments over each outcome across all the bandwidths used. The dashed line represents the confidence interval of 95 percent. As shown by the models, it is not possible to differentiate the effect of receiving the incentives from zero at this confidence level for schools near the cutoff.

[FIGURE 7]

[FIGURE 8]

[FIGURE 9]

[FIGURE 10]

5.2 Assessing heterogenous effects

Table 5 shows the results of the linear model controlled for pretreatment characteristics for the effect of incentive 1 on scores 2011 and on school value-added for different subsamples. The rationale for studying different subsamples of schools is that the amount of the incentive might not represent the same impact for a low socioeconomic school than for a school with relatively better resources. Since schools in more vulnerable conditions are easily identified for being located in small, rural and indigenous communities, it is possible to restrict the population to study them separately. It is expected that the bonuses might have a positive and significant effect on scores and value-added in a school with relatively few resources because that money would represent a larger proportion of the teachers' salary. This question is particularly relevant because for these schools the incentives would be multiplied by a factor of 1.1 in order to provide them with more support than regular schools.

The subsamples presented in table 5 are the different types of primary schools (regular, CONAFE and indigenous) and secondary schools (regular, CONAFE and technical/distance learning). I also identified samples for schools in rural and poor areas. In general, there is no significant effect of receiving the incentive for having a high score compared to the schools in the cluster that did not receive the incentive. However, the effect on school value-added is negative and significant at 95 percent for primary indigenous schools. For these schools, the effect of receiving High Achieving incentive reduced their change in scores from 2010 to 2011 in 0.2 units. In perspective, this figure represents one quarter of standard deviation of the indigenous schools' value-added. The standard deviation from the random model is low enough to consider that the effect is negative across different clusters.

[TABLE 5]

Table 6 confirms the results of the previous table. Among the subsamples studied, only indigenous schools show a negative and significant effect across the bandwidths. In this case, the effect is negative and significant for the 2011 ENLACE score. The effect is also close to a quarter of a standard deviation. Again, the results of the random model show that this effect is negative across clusters.

[TABLE 6]

5.3 Effects of receiving two awards

Table 7 shows the results for the RSRD and for the FRD to assess the effect of receiving two incentives on 2011 test score and on school value-added. These models include two rating scores and pre-treatment variables to increase precision. According to the RSRD, receiving both incentives does not show a significant effect on the outcomes studied. Using an FRD to study the effect of receiving two incentives compared to only receiving incentive 2 also shows no significant effect on 2011 test score nor on school value-added. On the contrary, the effect of receiving two incentives compared to only receiving incentive 1 shows a negative and significant impact on score 2011 for bandwidths of 25, 20 and 15 points from the cutoff. The variance of the coefficient across clusters show that in some cases the effect could be positive, although for the bandwidth of 15 it is small enough to be negative across all clusters. The size of this effect is close to a tenth of a standard deviation for score 2011. For school value-added there is no effect for this specification of FRD.

[TABLE 7]

6 Discussion

The results presented above show that, in general, there is no significant effect of receiving either of these incentives on the outcome variables. However, this effect seems to be negatively affecting indigenous primary schools. The explanations for the lack of effect of this incentive can be related to similar findings in other programs with no significant effect. In addition, the PECD seems to have a particular design problem because the first round was implemented more as a reward than an incentive. That is, prior to the implementation of the program, schools were not aware of the possibility of this incentive and thus there was not a process of building expectations that might be associated with increasing performance.

Another possibility is that the amount of the incentives was not enough to modify behavior. In Mexico the average yearly salary for a teacher in primary public schools is around 6,600 dollars and 7,400 for teachers in secondary schools (Santibañez 2002). The average incentive distributed by the PECD was about 280 dollars, which is close to four percent of yearly salary. This percentage is actually very close to the amount distributed by other incentive programs around the world. For example, a program piloted in New York City schools in 2007 with no significant effect on achievement distributed incentives close to 4.1 percent of average annual teacher salary (Fryer 2011). However, the amount of the incentive might not play a significant role since more successful programs distributed even smaller amounts in India (three percent) and Kenya (two percent). Moreover, a program distributing a large percentage of teachers' salary in Nashville, Tennessee, (up to 22 percent) had no significant effect on student achievement (Springer et al. 2010).

Another plausible explanation is that there has not been enough time for the program to have a positive effect. Due to data availability, this paper only studies the effect of the first round of incentives of the PECD, one year after they were distributed. First, this could mean that schools and teachers have not had time to adjust their behavior and completely understand the nature of this program. It is likely that teachers in schools receiving the incentives did not consider this to be a long-term program where they could continue to receive these payments as long as their students continued to outperform other schools. Second, the program might have a delayed effect that is not present yet. In this case, it would be worth it to study the effect of the incentives in the ENLACE scores of 2012 (two years after the implementation) or even in 2013 (three years after). It is not unlikely that the incentive is changing other behaviors that are not captured by standardized tests, like teacher attendance or job satisfaction, which might have an effect in students' long-term academic career not captured by the following year's test. As this program continues to be implemented with some adjustments and increasing budget, it is possible that cumulative effects could exist at schools receiving incentives during several years that are not yet identified.

Another classical argument for the non-impact of incentives schemes is the inability of teachers to improve students' performance with no further resources than improved salary. In this case, even if they feel more motivated by the incentives, they may not know how to raise student achievement on the standardized test. A perennial problem in the Mexican educational system is the poor quality of instruction and poor pre-service and in-service training of teachers (Santibañez 2004). About 40 percent of teachers in Mexico have never attended a teacher education institution, and those who have only spent about 15 percent of their training time learning subject matter (Santibañez et al. 2005). In this scenario, there could be little knowledge of how to improve student achievement even with a monetary incentive.

Here is important to consider the nature of the evaluation implemented since only studies schools that are near the cutoff. Although this provides a valid counterfactual, it also limits the analysis for a particular set of schools. In this case, it is likely schools that were just below the cutoff increased their supply of labor. Then, even if schools just above the cutoff were not doing worst than the previous year they could have been overpassed by the schools just above the cutoff, thus generating an overall negative effect of receiving the award.

Finally, a more difficult question raised by this analysis is why the incentive might be negatively affecting performance in indigenous schools. Along with CONAFE schools, these schools have the largest percentages of teachers who have only middle-school studies (OECD 2011); they also have important disadvantages in basic infrastructure and the socioeconomic level of their students. However, CONAFE schools do not present such effects. In this case, the differentiated effects might be related to one unique characteristic of those schools: their orientation to bilingual population in indigenous communities. Potential explanations might include the possibility that such remote communities did not receive the incentive they were assigned, an unanticipated cultural perception of monetary incentives, or a demoralizing effect of this kind of program that might be reducing effort (Fehr and Schmidt 2004). However more research is needed in order to fully understand the causes of such impact.

7 Conclusion

This paper provides the first impact evaluation of the Mexican PECD. It builds on the literature analyzing teacher and school incentives based on student performance and in the scarce subset of literature analyzing national-level incentive programs. I find that there is no evidence of an effect of the PECD on student achievement. If any, this program seems to have a negative effect; this effect is significant for indigenous primary schools that are affected for a 25 percent of a standard deviation on ENLACE score 2011 and on school value-added 2010-2011 for receiving the incentive. These results provide evidence of the difficulty of designing and implementing economic incentives in education.

The results of this study should be taken cautiously. First, this research only analyzes the first round of the PECD, thus there might be a learning curve for teachers and school administrators. This point is particularly relevant because the first application of the PECD can be characterized more as a reward than an incentive. Secondly, there is a question of generalization attached to the techniques employed here. The results presented here apply only for schools near the cutoff. Given that there is a large number of cutoffs, it is at least possible to say that they are generalizable at least for a range of 15,000-40,000 schools , which represent around 30 percent and 10 percent of all eligible Mexican schools. These results can also be generalized for schools near the cutoff within particular subgroups. Exploring the effects of the program in subsequent years and the effects of the incentives delivered directly to teachers (that accounted for 13 percent in the first of the first PECD round) is another avenue for future research. An additional topic of interest is related to the “noise” in the results of standardized tests that may limit their use in school rankings and in accountability systems (Kane and Staiger 2002; Chay, McEwan and Urquiola 2003).

Finally, this research leaves open important questions about the perception of teachers of these incentives and how they affect their behavior and effort they put in their job. A general policy recommendation for this program is that it should be attached to the provision of training schemes to provide teachers with further elements to increase student performance.

8 References

Amador Hernández, Juan Carlos. *La Alianza por la Calidad Educativa: Modernización de los centros escolares y profesionalización de los maestros.* Mexico City: Centro de Estudios Sociales y de Opinión Pública, 2009.

Angrist, Joshua, and Victor Lavy. "The Effects of High Staes High-School Achievement: Evidence from a Randomized Trial." *American Economic Review* 80, no. 3 (2009): 313-36.

Atkinson, Adele; Burgess, Simon; Croxson, Bronwyn; Paul, Gregg; Popper, Carol; Slater, Helen; Wilson, Deborah. "Evaluating the Impact of Perfomance-related Pay for Teachers in England." *Labour Economics* (16) 3 (2009): 251-261.

Ballou, Dale. "Pay for Performance in Public and Private Schools." *Economics of Education Review* 20 (2001): 51-61.

Campbell, Donald T., and Julian Stanley. *Experimental and Quasi-experimental Design for Research on Teaching.* Chicago: Cengage Learning, 1963.

Chay, Kenneth, Patrick McEwan, and M. Urquiola. "The Central Role of Noise in Evaluating Interventions that Use Test Scores to Rank Schools." *NBER Working Paper No. 10118*, 2003.

Cook, Thomas D. "'Waiting for Life to Arrive': A history of the regression-discontinuity design in Psychology, Statistics and Economics." *Journal of Econometrics*, 2007: 636-654.

Dixit, Avinsah. "Incentives and Organizations in the Public Sector. An Interpretative Review." *The Journal of Human Resources* 37, no. 7 (2002): 696-727.

Duflo, Esther, Rema Hanna, and Stephen Ryan. "Monitoring Works: Getting Teachers to Come to School." *American Economic Review* 102, no. 4 (2012): 1241-78.

Fehr, Ernst, and Klaus M. Schmidt. "Fairness and Incentives in a Multi-Task Principal-Agent Model." *The Scandinavian Journal of Economics* 106, no. 4 (2004): 453-74.

Figlio, David, and Lawrence Kenny. "Individual teacher incentives and student performance." *Journal of Public Economics* 91 (2007): 901-14.

Fryer, Roland G. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *The Quartelry Journal of Economics* 126, no. 4 (2011): 1755-98.

Fryer, Roland. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." *NBER Working Paper No. 16850*, 2011.

Glewwe, Paul, Michael Kremer, and Ilias Nauman. "Teacher Incentives." *American Economic Journal: Applied Economics* 2, no. 3 (2010): 205-227.

Gonring, Phil. *Pay-for-performance teacher compensation : an inside view of Denver's ProComp plan.* Cambridge, Mass.: Harvard Education Press, 2007.

Goodman, Sarena, y Lesley Turner. «Teacher Incentive Pay and Educational Outcomes: Evidence from the NYC Bonus Program.» *Program on Education Policy and Governance Working Papers Series*, 2010.

Holmstrom, Bengt, and Paul Milgrom. "Multitask Principal-Agent Analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics & Organization* 7 (1991): 24-52.

Hoxby, Caroline, and Andrew Leigh. "Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States." *The American Economic Review* 94, no. 2 (2004): 236-240.

Hudson, Sally. "The Effects of Performance-Based Teacher Pay on Student Achievement." *SIEPR Discussion Paper no. 09-023*, 2010.

Imbens, Guido, and Tristan Zajoc. "Regression Discontinuity Design with Multiple Forcing Variables." *Working paper*, 2011.

Instituto Nacional de Evaluación Educativa (INEE). *El Derecho a la Educación en México, Informe 2009.* INEE, 2010.

Kane, Thomas J. "A Quasi-Experimental Estimate of the Impact of Financial Aid on College-Going." *NBER Working Paper No. 9703*, 2003.

Kane, Thomas, and Douglas Staiger. "Improving School Accountability Measures." *NBER Working Paper No. 8156*, 2001.

Koretz, Daniel. "Limitations in the Use of Achievement test as Measures of Educator's Productivity." *Journal of Human Resources* 37, no. 4 (2002): 752-777.

Koretz, Daniel, and Sheila I. Barron. *The Validity of Gains on the Kentucky Instructional Results Information System (KIRIS).* Santa Monica: RAND, 1998.

La Jornada. *Repudian en Michoacán y Oaxaca alianza educativa pactada por Calderón y Gordillo.* September 13, 2008. http://www.jornada.unam.mx/2008/09/13/index.php?section=sociedad&article=041n1so.

Ladd, Hellen F. "The Dallas School Accountability and Incentive Program: an Evaluation of its Impact on Student Outcomes." *Economic of Education Review*, 1999: 1-16.

Lavy, Victor. "Evaluating the Effect of Teacher Group Performance Incentives on Pupil Achievements." *The Journal of Political economy* 110, no. 6 (2002): 1286-1318.

Lavy, Victor. "Using Performance-Based Pay to Improve the Quality of Teachers." *The Future of Children* 17, no. 1 (2007): 87-109.

Lee, David S. "Randomized Experiments from Non-random Selection in U.S. House Elections." *Journal of Econometrics* 142, no. 2 (2008): 675-697.

Martins, Pedro. "Individual Teacher Incentives, Student Achievement and Grade Inflation." *IZA Discussion Paper No. 4051*, 2009.

Mizala, Alejandra, and Pilar Romanguera. *Teachers' Salary Structure and Incentives in Chile.* Center for Applied Economics, Department of Industrial Engineering, University of Chile, 2004.

Muralidharan, Khartik, and Venkatesh Sundararaman. "Teacher performance pay: Experimental evidence from India." *Research working paper* (Journal of Political Economy) 119, no. 1 (2011): 39-77.

Neal, Derek. "The Design of Performance Pay in Education." *NBER Working Paper No. 16710*, 2011.

Organization for Economic Cooperation and Development (OECD). *Establishing a Framework for Evaluation and Teacher Incentives. Considerations for Mexico.* Paris: OECD, 2011.

Pendergast, Canice. "The Provision of Incentives in Firms." *Journal of Economic Literature*, 1999: 7-63.

Raundenbush, Stephen W., and Anthony S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Thoudand Oaks, CA: Sage, 2002.

Reardon, Sean F., Allison Atteberry, Nicole Arshan, and Michal Kurlaender. "Effects of the California High School Exit Exam on Student Persistence, Achievement, and Graduation." *Educational Evaluation and Policy Anlysis* 32, no. 4 (2010): 498-520.

Robinson, Joseph P., and Sean F. Reardon. "Regression Discontinuity Designs with Multiple Rating-Score Variables." *Working paper*, 2010.

Santibañez, Lucrecia. "¿Están Mal Pagados los Maestros en México? Estimado de los Salarios Relativos del Magisterio." *Revista Latinoamericana de Estudios Educativos* 32, nº 2 (2002): 9-41.

Santibañez, Lucrecia, Geroges Vernez, and Razquin Paula. *Education in Mexico. Challenges and Opportunities.* Santa Monica: RAND, 2005.

Santibañez, Lucrecia, y otros. *Haciendo Camino. Análisis del sistema de evaluación y del programa de estímulos docentes Carrera Magisterial en México.* Santa Monica: RAND, 2006.

Secretaría de Educación Pública (SEP). *Acciones para la Entrega del Apoyo Financianciero del Programa de Estímulos a la Calidad Docente.* SEP, 2010.

Secretaría de educación Pública (SEP). *Memoria Técnica del Método de Asignación de Recursos del Programa de Estímulos a la Calidad Docente.* Mexico City: SEP, 2010.

Springer, Matthew; Ballou, Dale; Hamilton, Laura; Vi-Nhuan, Le; Lockwood, J.R.; McCaffrey, Daniel; Pepper, Matthew; Stecher, Brian . *Teacher Pay for Performance. Experimenal Evidence from the Project on Incentives in teaching.* Nashville: The National Center on Performance Incentives, 2010.

Thistlethwaite, Donald L., and Donald T. Campbell. "Regression-discontinuity Analysis: An alternative to the ex-post facto experiment." *Journal of Educational Phsychology* 51 (1978): 309-317.

Umansky, Ilana. "A Literature Review of Teacher Quality and Incentives. Theory and Evidence." In *Incentives to Improve Teachers. Lessons from Latin America*, by The World Bank, edited by E. Vegas and I. Umansky, 21-62. Washington, DC: The World Bank, 2005.

Vigdor, Jacob L. "Teacher Salary Bonuses in North Carolina." In *Perfromance Incentives: Their Growing Impact on American K-12 Education*, edited by Matthew Springer. Brookings, 2009.

Winters, Marcus, Jay Greene, Gary Ritter, and Ryan Marsh. *The Effect of Performance-Pay in Little Rock, Arkansas on Student Achievement.* National Center in Performance Incentives, Peabody College of Vanderbilt University, 2008.

9 Figures

Figure 1. Cluster criteria for PECD assignation in each of Mexican State

	Primary			Secondary			
	General	**Communitary**	**Indigenous**	**General***	**Communitary**	**Technical**	**Distance learning (Telesecundaria)**
URBAN Rich							
Poor							
RURAL Rich							
Poor							

Note: Due to the small amount of workers oriented secondary schools these schools are merged with general secondary schools. The smallest number of observations allowed for each of these clusters was 10 schools. In case there were less than 10 schools, the cluster is merged in the following order until it completes the minimum required number: by type of locality, by vulnerability level and, finally, by type of service.
Source: Author with data from Memoria Técnica del Método de Asignación de Recursos del Programa de Estímulos a la Calidad Docente (SEP 2010).

Table 1. Schools Benefited by the PECD in each State and Number of Clusters

STATE	High achieving (T1)			Most improving (T2)			T1 AND T2			Clusters
	NO	YES	Total	NO	YES	Total	NO	YES	Total	
AGUASCALIENTES	810	137	947	809	138	947	919	28	947	10
	85.53	*14.47*	*100*	*85.43*	*14.57*	*100*	*97.04*	*2.96*	*100*	
BAJA CALIFORNIA	1,530	259	1,789	1,530	259	1,789	1,749	40	1,789	17
	85.52	*14.48*	*100*	*85.52*	*14.48*	*100*	*97.76*	*2.24*	*100*	
BAJA CALIFORNIA SUR	460	68	528	455	73	528	515	13	528	10
	87.12	*12.88*	*100*	*86.17*	*13.83*	*100*	*97.54*	*2.46*	*100*	
CAMPECHE	911	141	1,052	904	148	1,052	1,019	33	1,052	15
	86.6	*13.4*	*100*	*85.93*	*14.07*	*100*	*96.86*	*3.14*	*100*	
COAHUILA	9,334	1,390	10,724	9,293	1,431	10,724	10,383	341	10,724	23
	87.04	*12.96*	*100*	*86.66*	*13.34*	*100*	*96.82*	*3.18*	*100*	
COLIMA	2,993	446	3,439	2,983	456	3,439	3,358	81	3,439	18
	87.03	*12.97*	*100*	*86.74*	*13.26*	*100*	*97.64*	*2.36*	*100*	
CHIAPAS	1,859	299	2,158	1,856	302	2,158	2,099	59	2,158	14
	86.14	*13.86*	*100*	*86.01*	*13.99*	*100*	*97.27*	*2.73*	*100*	
CHIHUAHUA	509	83	592	508	84	592	573	19	592	10
	85.98	*14.02*	*100*	*85.81*	*14.19*	*100*	*96.79*	*3.21*	*100*	
DISTRITO FEDERAL	2,678	461	3,139	2,674	465	3,139	3,045	94	3,139	7
	85.31	*14.69*	*100*	*85.19*	*14.81*	*100*	*97.01*	*2.99*	*100*	
DURANGO	3,010	457	3,467	3,004	463	3,467	3,370	97	3,467	16
	86.82	*13.18*	*100*	*86.65*	*13.35*	*100*	*97.2*	*2.8*	*100*	
GUANAJUATO	5,004	837	5,841	4,994	847	5,841	5,677	164	5,841	17
	85.67	*14.33*	*100*	*85.5*	*14.5*	*100*	*97.19*	*2.81*	*100*	
GUERRERO	5,780	590	6,370	5,540	830	6,370	6,220	150	6,370	24
	90.74	*9.26*	*100*	*86.97*	*13.03*	*100*	*97.65*	*2.35*	*100*	
HIDALGO	3,634	598	4,232	3,627	605	4,232	4,142	90	4,232	21
	85.87	*14.13*	*100*	*85.7*	*14.3*	*100*	*97.87*	*2.13*	*100*	
JALISCO	6,137	980	7,117	6,133	984	7,117	6,938	179	7,117	19
	86.23	*13.77*	*100*	*86.17*	*13.83*	*100*	*97.48*	*2.52*	*100*	
MÉXICO	8,458	1,410	9,868	8,452	1,416	9,868	9,639	229	9,868	25
	85.71	*14.29*	*100*	*85.65*	*14.35*	*100*	*97.68*	*2.32*	*100*	
MICHOACAN	4,672	220	4,892	4,653	239	4,892	4,865	27	4,892	16
	95.5	*4.5*	*100*	*95.11*	*4.89*	*100*	*99.45*	*0.55*	*100*	
MORELOS	1,009	161	1,170	993	177	1,170	1,140	30	1,170	16
	86.24	*13.76*	*100*	*84.87*	*15.13*	*100*	*97.44*	*2.56*	*100*	
NAYARIT	1,443	235	1,678	1,438	240	1,678	1,623	55	1,678	17
	86	*14*	*100*	*85.7*	*14.3*	*100*	*96.72*	*3.28*	*100*	
NUEVO LEÓN	2,795	443	3,238	2,790	448	3,238	3,154	84	3,238	14
	86.32	*13.68*	*100*	*86.16*	*13.84*	*100*	*97.41*	*2.59*	*100*	
OAXACA	7,289	181	7,470	7,271	199	7,470	7,427	43	7,470	23
	97.58	*2.42*	*100*	*97.34*	*2.66*	*100*	*99.42*	*0.58*	*100*	
PUEBLA	5,268	712	5,980	5,113	867	5,980	5,825	155	5,980	21
	88.09	*11.91*	*100*	*85.5*	*14.5*	*100*	*97.41*	*2.59*	*100*	
QUERÉTARO	1,421	244	1,665	1,421	244	1,665	1,616	49	1,665	15
	85.35	*14.65*	*100*	*85.35*	*14.65*	*100*	*97.06*	*2.94*	*100*	
QUINTANA ROO	830	136	966	825	141	966	937	29	966	13
	85.92	*14.08*	*100*	*85.4*	*14.6*	*100*	*97*	*3*	*100*	
SAN LUIS POTOSI	4,226	682	4,908	4,217	691	4,908	4,798	110	4,908	19
	86.1	*13.9*	*100*	*85.92*	*14.08*	*100*	*97.76*	*2.24*	*100*	
SINALOA	3,100	454	3,554	3,092	462	3,554	3,482	72	3,554	17
	87.23	*12.77*	*100*	*87*	*13*	*100*	*97.97*	*2.03*	*100*	
SONORA	2,014	315	2,329	2,010	319	2,329	2,269	60	2,329	19
	86.47	13.53	100	86.3	13.7	100	97.42	2.58	100	
TABASCO	2,349	385	2,734	2,346	388	2,734	2,629	105	2,734	18
	85.92	*14.08*	*100*	*85.81*	*14.19*	*100*	*96.16*	*3.84*	*100*	
TAMAULIPAS	2,557	415	2,972	2,550	422	2,972	2,894	78	2,972	16
	86.04	*13.96*	*100*	*85.8*	*14.2*	*100*	*97.38*	*2.62*	*100*	
TLAXCALA	820	144	964	818	146	964	936	28	964	17
	85.06	*14.94*	*100*	*84.85*	*15.15*	*100*	*97.1*	*2.9*	*100*	
VERACRUZ	10,638	1,729	12,367	10,631	1,736	12,367	12,047	320	12,367	23
	86.02	*13.98*	*100*	*85.96*	*14.04*	*100*	*97.41*	*2.59*	*100*	
YUCATAN	1,468	251	1,719	1,468	251	1,719	1,652	67	1,719	18
	85.4	*14.6*	*100*	*85.4*	*14.6*	*100*	*96.1*	*3.9*	*100*	
ZACATECAS	2,756	442	3,198	2,747	451	3,198	3,106	92	3,198	16
	86.18	*13.82*	*100*	*85.9*	*14.1*	*100*	*97.12*	*2.88*	*100*	
TOTAL	**107,762**	**15,305**	**123,067**	**107,145**	**15,922**	**123,067**	**120,046**	**3,021**	**123,067**	**544**
	87.56	**12.44**	**100**	**87.06**	**12.94**	**100**	**97.55**	**2.45**	**100**	

Notes: First row for each State represents the number of schools in each category; second row represents the percentage of schools. Number of clusters represents the number of subdivisions implemented by Instituto Nacional para la Evaluación de la Educación (INEE) showed in Figure 1.
Source: Author with data from Programa de Estímulos a la Calidad Docente (http://estimulosalianza.sep.gob.mx/?p=reconocidos) and with data retrieved via Instituto Federal de Acceso a la Información (IFAI).

Table 2. Indexes used as pre-treatment controls and variables included

Index	Questions included
Basic infrastructure index	Does the school have: Drinkable water? Sewage? Electric light? Air conditioning? Garbage collection? Cleaning services? Satellite? Internet? Telephone? School transportation? Cafeteria?
School infrastructure index	Are the following facilities adequate for the school: Classrooms? Laboratories? Workshops? Library? Administrative spaces? Recreational spaces? Auditorium? Computer cluster? Teachers' lounge? Bathrooms? Facilities for disabled people?
Educational material index	Are the following facilities adequate for the school: computers for the students? Computers for the teachers? Computers for administrators? Electric blackboard? Media equipment? Laboratory equipment? Equipment for the workshops? Furniture for the students? Furniture for disabled people? Materials for class activities?
Personnel index	Beside teachers, does the school have: Assistant principal? Counselor? Psychologists? Prefect? Janitor? Supporting staff? Elective teachers? Bilingual teachers? Bilingual teachers that speak the native language of the students (applies for indigenous schools)?
Teacher quality index	Frequency in which teachers: Arrive late? Do not attend classes? Do not complete scheduled time for class? Cannot control the group? Have low performance? Have insufficient computational skills? Resist professional development? Resist change of grade/subject? Disregard curriculum?
Evaluation index	Frequency in which teachers are evaluated in: Attendance? Punctuality? Seniority? Mastery of content? Teaching techniques? Professional development? Planning? Academic support for students? Evaluation methods of learning? Student achievement? Disciplinary control of the group? Participation in academic activities?
Dropout index	Which of the following activities is a cause of student dropout: Low performance? Lack of interest in school? Disciplinary problems? Lack of support from parents? Health problems? Economic problems? Migration? Addiction? Pregnancy? Safety problems around the school? Bad relationships among the students? Bad relationships with teachers? Bad relationships with school authorities?
Internal environment index	Frequency of the following situations in the school: Bullying? Violence towards teachers? Theft? Fights? Bearing of weapons? Consumption of alcoholic beverages? Consumption of drugs?
External environment index	Frequency of the following situations in the school surroundings: Lack of street lights? Lack of public surveillance? Drug consumption? Robbery/assault? Gang activity? Vandalism? Sale of alcoholic beverages?

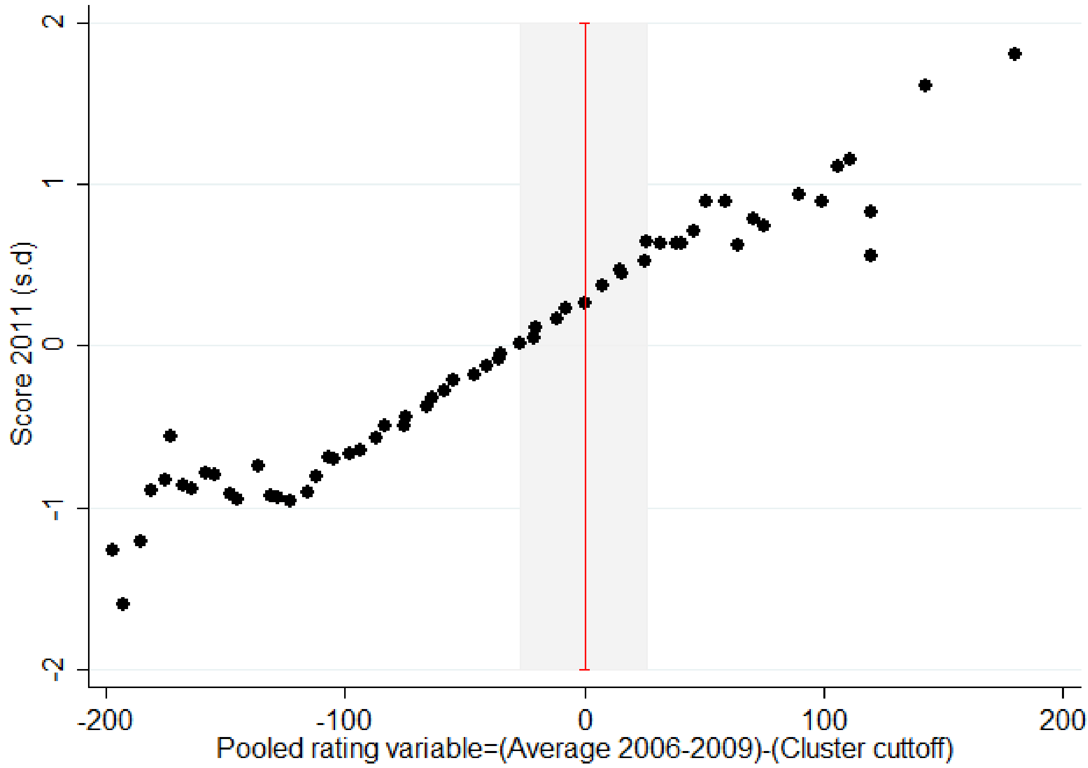
Notes: All indexes where created using a principal components analysis and then predicting the new variable using the varimax rotated factors which was then standardized from 0 to 1.

Table 3. Pre-treatment Characteristics Balance Check

	Primary schools (general)*						Secondary schools (general)					
	High achieving incentive (T1)			Best-improving incentive (T2)			High achieving incentive (T1)			Best-improving incentive (T2)		
	T1=0	T1=1		T2=0	T2=1		T1=0	T1=1		T2=0	T2=1	
	mean	mean	p-value	mean	mean	p-value	mean	mean	p-value	mean	mean	p-value
Total number of students	54.94	54.31	0.77	54.24	50.32	0.09	379.48	427.69	0.00	358.63	322.63	0.00
Basic infrastructure index	0.33	0.33	0.60	0.32	0.31	0.03	0.70	0.72	0.02	0.71	0.69	0.01
School infrastructure index	0.27	0.29	0.01	0.27	0.27	0.88	0.63	0.66	0.01	0.64	0.62	0.06
Educational material index	0.31	0.32	0.05	0.32	0.32	0.76	0.52	0.55	0.02	0.52	0.51	0.23
Personnel index	0.04	0.04	0.52	0.04	0.04	0.07	0.54	0.55	0.15	0.54	0.53	0.27
Teacher quality index	0.86	0.87	0.15	0.86	0.87	0.11	0.68	0.69	0.67	0.67	0.67	0.97
Evaluation index	0.59	0.59	0.99	0.59	0.59	0.29	0.58	0.59	0.19	0.58	0.58	0.85
Droput index	0.12	0.11	0.68	0.12	0.12	0.26	0.26	0.25	0.29	0.29	0.28	0.27
Internal environment index	0.06	0.06	0.12	0.06	0.06	0.15	0.17	0.18	0.31	0.19	0.18	0.36
External environment index	0.15	0.14	0.02	0.14	0.14	0.86	0.33	0.33	0.76	0.35	0.34	0.26
	Primary schools (CONAFE)						**Secondary schools (CONAFE)**					
Total number of students	5.90	6.15	0.49	6.26	6.14	0.74	11.59	11.65	0.98	12.46	9.91	0.08
Basic infrastructure index	0.20	0.23	0.03	0.20	0.21	0.59	0.21	0.29	0.01	0.24	0.24	0.94
School infrastructure index	0.21	0.22	0.17	0.22	0.21	0.53	0.21	0.22	0.69	0.21	0.24	0.43
Educational material index	0.21	0.22	0.16	0.22	0.21	0.62	0.29	0.32	0.34	0.31	0.33	0.65
Personnel index	0.03	0.03	0.49	0.03	0.03	0.74	0.02	0.01	0.40	0.04	0.03	0.73
Teacher quality index	0.88	0.89	0.38	0.88	0.87	0.71	0.90	0.90	0.97	0.87	0.86	0.81
Evaluation index	0.50	0.51	0.53	0.52	0.52	0.56	0.53	0.52	0.75	0.56	0.57	0.80
Droput index	0.16	0.15	0.45	0.17	0.18	0.47	0.16	0.21	0.17	0.18	0.19	0.89
Internal environment index	0.04	0.03	0.02	0.04	0.05	0.06	0.03	0.03	0.98	0.06	0.03	0.16
External environment index	0.08	0.08	0.92	0.07	0.09	0.01	0.10	0.13	0.21	0.12	0.08	0.29
	Primary schools (indigenous)						**Secondary schools (Distance learning -Telesecundaria)**					
Total number of students	60.29	56.39	0.42	52.82	53.16	0.95	74.67	77.20	0.37	71.82	63.21	0.00
Basic infrastructure index	0.27	0.28	0.62	0.25	0.24	0.66	0.39	0.40	0.09	0.39	0.37	0.04
School infrastructure index	0.24	0.25	0.20	0.24	0.25	0.31	0.33	0.33	0.92	0.32	0.31	0.06
Educational material index	0.26	0.27	0.34	0.26	0.26	0.46	0.42	0.41	0.88	0.42	0.40	0.09
Personnel index	0.14	0.15	0.50	0.15	0.16	0.60	0.05	0.05	0.60	0.05	0.04	0.01
Teacher quality index	0.84	0.84	0.67	0.84	0.83	0.23	0.84	0.85	0.43	0.85	0.85	0.30
Evaluation index	0.59	0.59	0.79	0.59	0.60	0.93	0.59	0.59	0.51	0.59	0.59	0.27
Droput index	0.16	0.16	0.63	0.17	0.18	0.39	0.17	0.16	0.27	0.18	0.17	0.30
Internal environment index	0.05	0.05	0.89	0.05	0.06	0.13	0.08	0.08	0.34	0.08	0.07	0.05
External environment index	0.15	0.16	0.52	0.15	0.16	0.35	0.22	0.21	0.24	0.22	0.20	0.01

Notes: P-value from T-test performed with a bandwidth of 15 points from the cut-off for all variables. General primary schools' sample is restricted to schools in poor and rural areas.
Soruce: Author with data from Programa de Estímulos a la Calidad Docente and from ENLACE 2010 contextual questionnaires.

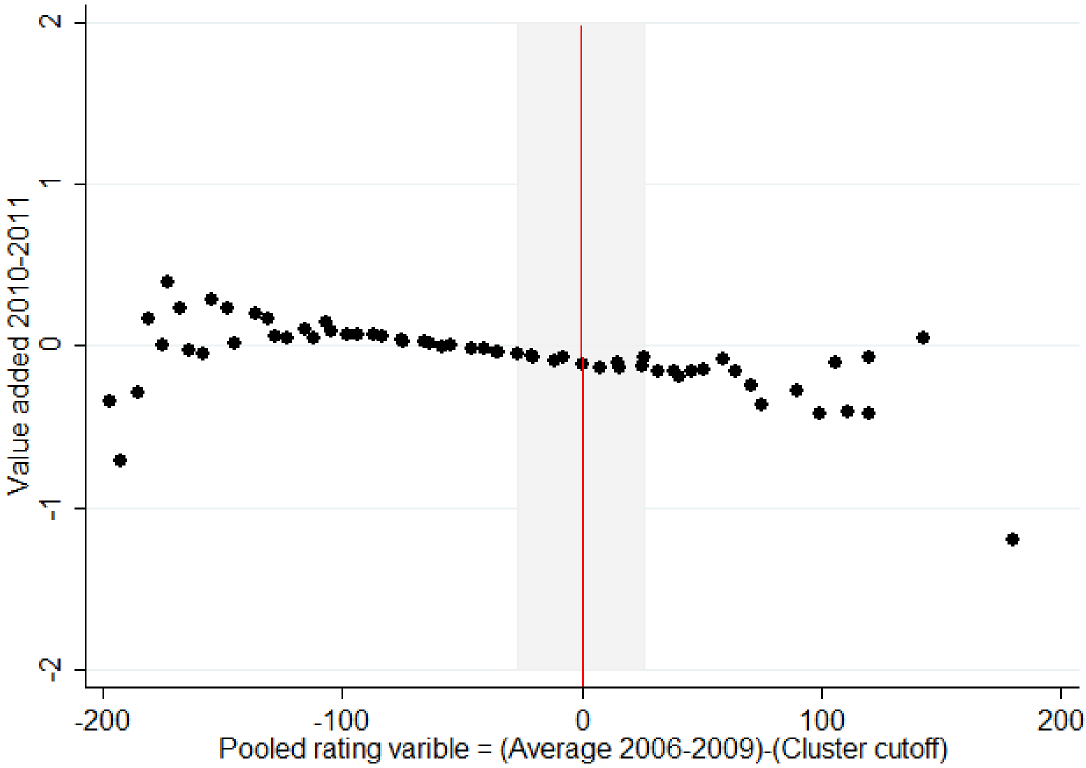
Figure 2. ENLACE test score 2011 and pooled rating score variable 1



Notes: Rating variable is the average of ENLACE scores 2006, 2007, 2008 and 2009 centered at zero for all clusters by subtracting the cluster cutoff. Outcome variable is the averages at 5 bin points of ENLACE 2011 standardized score. Shaded area represents a bandwidth of 30 points. Vertical line at zero represents the pooled frontier.

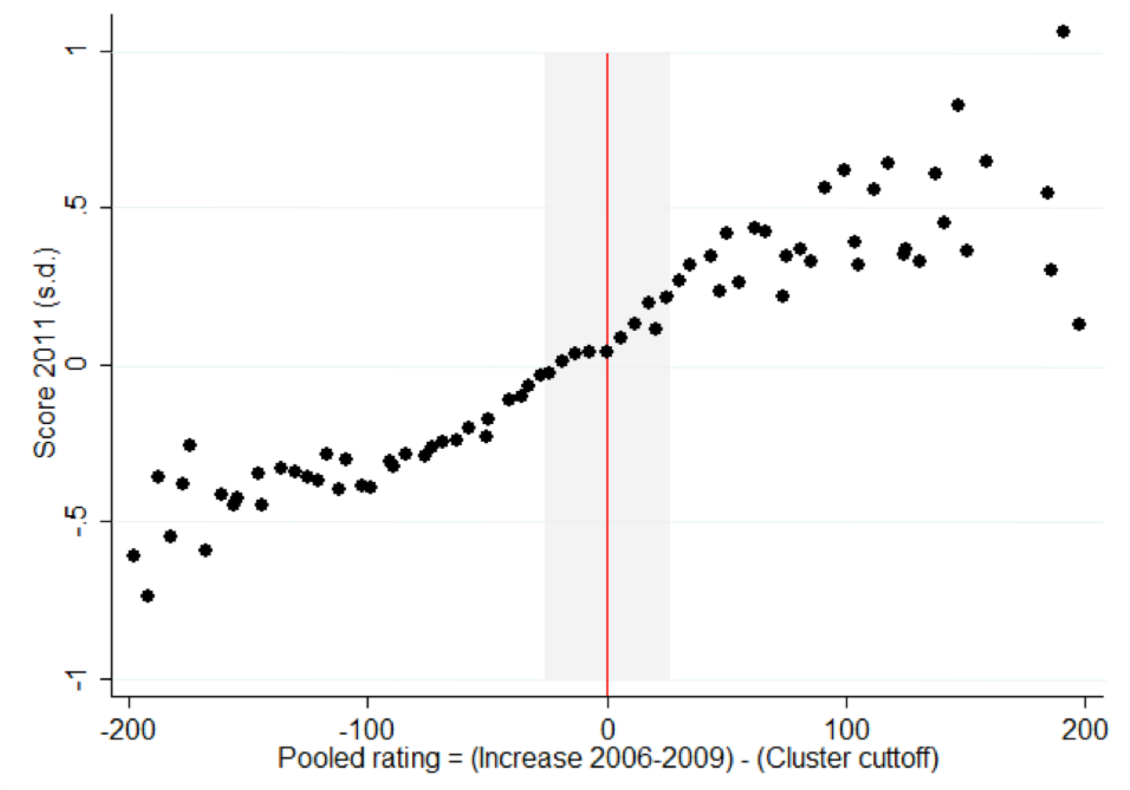
Source: Author with data from Programa de Estímulos a la Calidad Docente and from ENLACE 2011.

Figure 3. School value added and pooled rating score variable 1



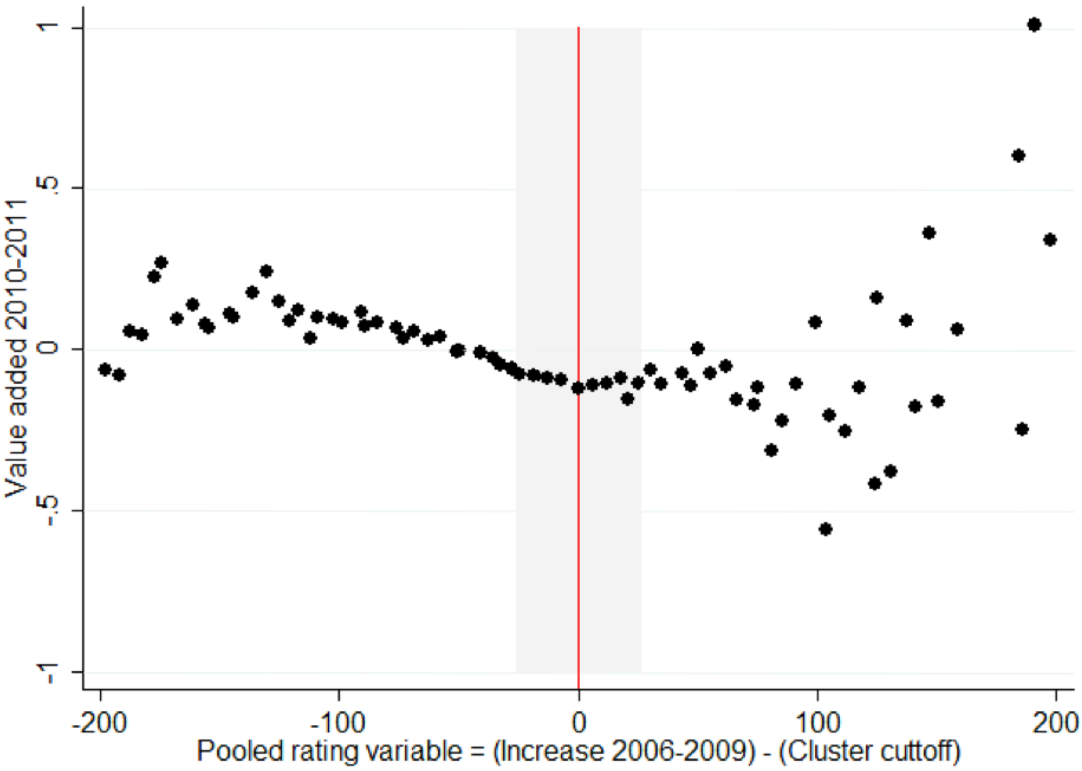
Notes: Rating variable is the difference between ENLACE 2009 and the average of all former years centered at zero for all clusters by subtracting the cluster cutoff. Outcome variable is the averages at 5 bin points of the change in the standardized scores between 2010 and 2011 ENLACE scores. Shaded area represents a bandwidth of 30 points. Vertical line at zero represents the pooled frontier.
Source: Author with data from Programa de Estímulos a la Calidad Docente and from ENLACE 2010 and 2011.

Figure 4. ENLACE test score 2011 and pooled rating score variable 2



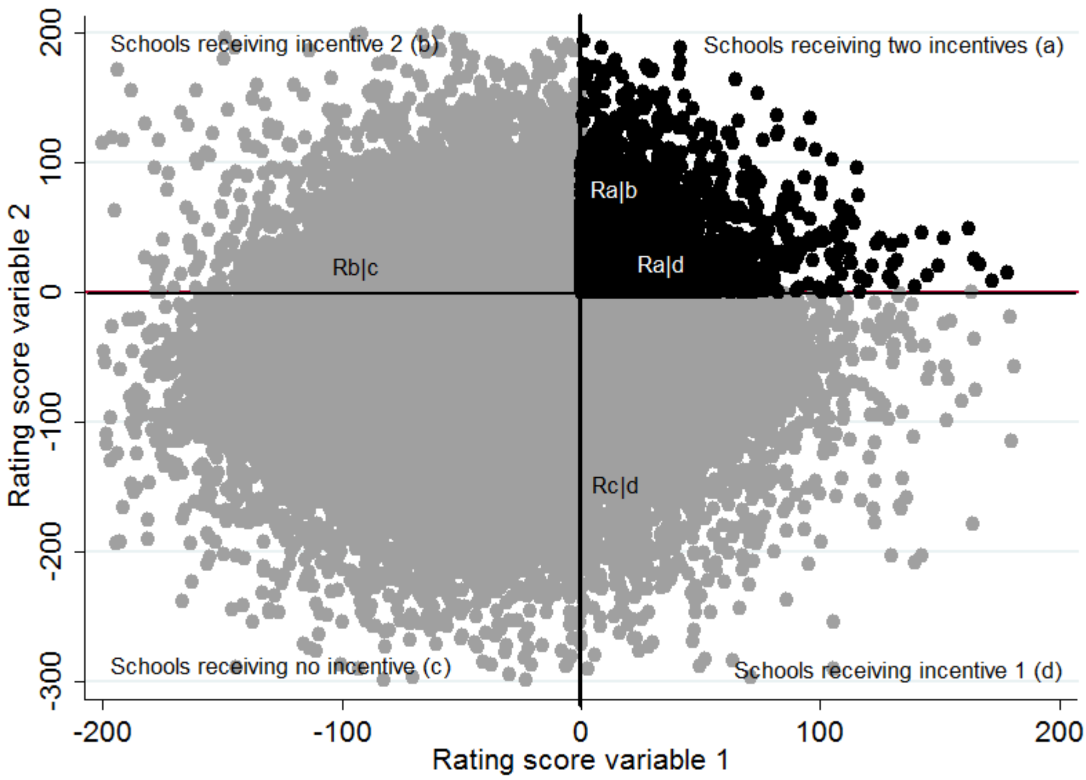
Notes: Rating variable is the average of ENLACE scores 2006, 2007, 2008 and 2009 centered at zero for all clusters by subtracting the cluster cutoff. Outcome variable is the averages at 5 bin points of ENLACE 2011 standardized score. Shaded area represents a bandwidth of 30 points. Vertical line at zero represents the pooled frontier.
Source: Author with data from Programa de Estímulos a la Calidad Docente and from ENLACE 2011.

Figure 5. School value added and pooled rating score variable 2



Notes: Rating variable is the difference between ENLACE 2009 and the average of all former years centered at zero for all clusters by subtracting the cluster cutoff. Outcome variable is the averages at 5 bin points of the change in the standardizes scores between 2010 and 2011 ENLACE scores. Shaded area represents a bandwidth of 30 points. Vertical line at zero represents the pooled frontier.
Source: Author with data from Programa de Estímulos a la Calidad Docente and from ENLACE 2010 and 2011.

Figure 6. Discontinuity with two rating score variables



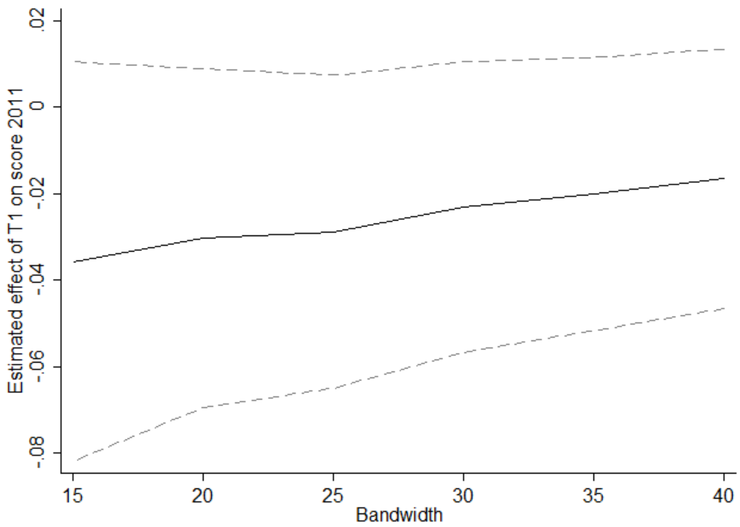
Notes: Pooled rating scores for High achievers and Best improving schools. Lines at zero represent the frontiers for each rating score. Quadrant a are schools that received two incentives; quadrant b are schools that received only incentive 2 (Best improving); quadrant c are schools that received no incentive and quadrant d are schools that received only incentive 1 (High achievers). Boundary Ra|b is the frontier between schools that received two incentives vs. only incentive2; boundary Rb|c is the frontier between schools that received incentive 2 vs. no incentive; boundary Rc|d represents the frontier between schools that received no incentive vs. only incentive1 and, boundary Ra|b is the frontier between schools that received only incentive1 vs. two incentives.
Source: Author with data from Programa de Estímulos a la Calidad Docente.

Table 4. Effect of each incentive on ENLACE test score 2011 and on school value-added 2010-2011

	Score 2011 (s.d.)						Value added 2010-2011					
	40	35	30	25	20	15	40	35	30	25	20	15
Linear model												
Effect of receiving incentive for high score (T1)	-0.0132	-0.0204	-0.0207	-0.0191	-0.0237	-0.0339	-0.00835	-0.00964	-0.0156	-0.0183	-0.0249	-0.0306
	(0.0138)	(0.0144)	(0.0151)	(0.0164)	(0.0180)	(0.0206)	(0.0123)	(0.0132)	(0.0140)	(0.0152)	(0.0169)	(0.0202)
Sd. (T1)	0.0756***	0.0713***	0.0590***	0.0592***	0.0570***	0.0526***	0.0342***	0.0393***	0.0340***	0.0337***	0.0388***	0.0764***
	(0.0123)	(0.0127)	(0.0140)	(0.0162)	(0.0192)	(0.0235)	(0.0163)	(0.0154)	(0.0166)	(0.0174)	(0.0196)	(0.0187)
n	46612	40835	34828	28876	22989	17184	46252	40520	34546	28640	22812	17059
Linear model with controls												
Effect of receiving incentive for high score (T1)	-0.0166	-0.0201	-0.0231	-0.0288	-0.0303	-0.0357	-0.0154	-0.0144	-0.0237	-0.0246	-0.0287	-0.0302
	(0.0150)	(0.0158)	(0.0168)	(0.0181)	(0.0196)	(0.0231)	(0.0135)	(0.0143)	(0.0154)	(0.0168)	(0.0186)	(0.0219)
Sd. (T1)	0.0681***	0.0674***	0.0629***	0.0582***	0.0344***	0.00181	0.0205**	0.0207***	0.0170***	0.0235***	0.0156	0.0454***
	(0.0133)	(0.0138)	(0.0154)	(0.0188)	(0.0301)	(0.0139)	(0.0243)	(0.0243)	(0.00385)	(0.0267)	(0.0477)	(0.0268)
n	35747	31306	26691	22124	17622	13219	35747	31306	26691	22124	17622	13219
Linear model												
Effect of receiving incentive for increase (T2)	-0.0294*	-0.0286~	-0.0273~	-0.0270~	-0.0310~	-0.0205	-0.00119	-0.00243	-0.000896	0.00157	-0.00439	0.000365
	(0.0145)	(0.0148)	(0.0155)	(0.0161)	(0.0174)	(0.0204)	(0.0125)	(0.0128)	(0.0140)	(0.0151)	(0.0160)	(0.0181)
Sd. (T2)	0.0836***	0.0737***	0.0634***	0.0237***	0.0243***	0.0139***	0.0777***	0.0718***	0.0921***	0.0884***	0.0910***	0.0848***
	(0.0149)	(0.0166)	(0.0187)	(0.00971)	(0.0104)	(0.0122)	(0.0142)	(0.0146)	(0.0154)	(0.0172)	(0.0178)	(0.0205)
n	51657	45983	39598	31524	26588	19030	51495	45843	39483	31436	26513	18968
Linear model with controls												
Effect of receiving incentive for increase	-0.00109	-0.00875	-0.00465	-0.00823	-0.0201	-0.0000783	0.00542	0.00220	0.0101	0.0126	0.00379	0.0104
	(0.0163)	(0.0167)	(0.0177)	(0.0193)	(0.0199)	(0.0229)	(0.0138)	(0.0141)	(0.0154)	(0.0163)	(0.0177)	(0.0200)
Sd. (T2)	0.101***	0.0935***	0.0929***	0.0954***	0.0669***	0.0456***	0.0751***	0.0669***	0.0865***	0.0752***	0.0879***	0.0788***
	(0.0156)	(0.0165)	(0.0171)	(0.0195)	(0.0220)	(0.0347)	(0.0168)	(0.0175)	(0.0171)	(0.0215)	(0.0218)	(0.0260)
n	39695	35361	30473	24274	20465	14609	39695	35361	30473	24274	20465	14609

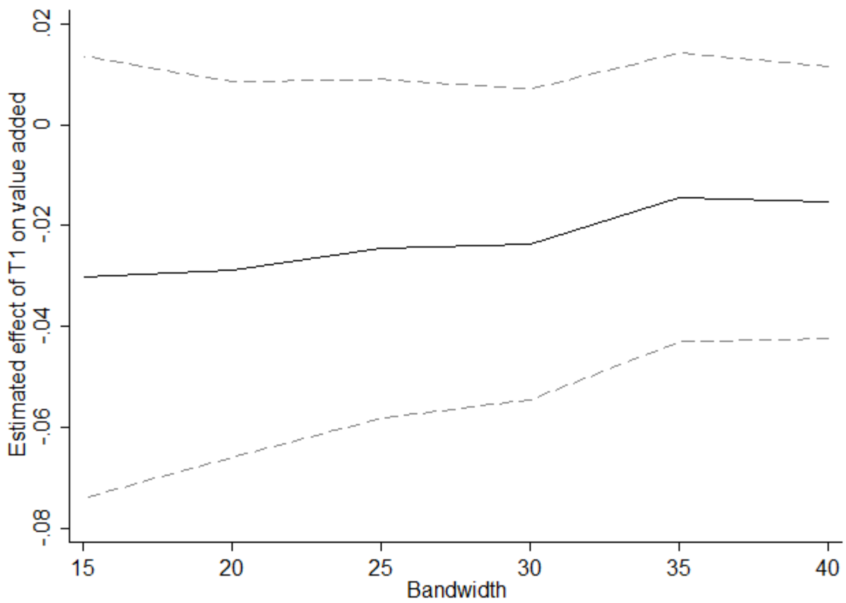
Notes: Score 2011 (s.d.) is the average ENLACE score for Mathematics and Spanish obtained for each school during 2011 standardized to have mean zero and variance equal to one. Value added 2010-2011 is the change in the standardized average ENLACE scores between 2010 and 2011. Standard errors are shown in parenthesis. The table reports the point estimate and the standard error on the coefficients of receiving each of the incentives as shown in model 2 in the text using a linear model with the same slope at each side of the cutoff. It also reports the standard deviation of the coefficient obtained from the random estimations of the mixed model. Covariates included in the second version of each model are: total number of students, basic infrastructure index, school infrastructure index, educational material index, personnel index, teacher quality index, evaluation index, dropout index, internal environment index and external environment index. All models restrict the sample to schools scoring within the bandwidth specified at the top of the column. All models exclude schools that scored above the cutoff but not received a reward. ~ $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Figure 7. Estimated effect of T1 on ENLACE score 2011 over different bandwidths



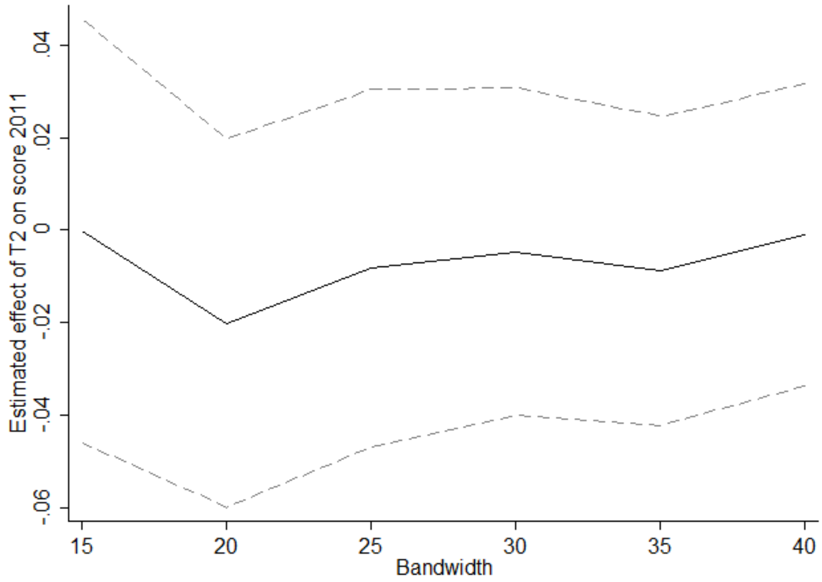
Notes: Solid lines represent estimated effects of receiving incentive 1 on ENLACE standardized score 2011 using model 2 in the text including covariates. Dashed lines represent 95% confidence intervals.

Figure 8. Estimated effect of T1 on school value-added 2010-2011 over different bandwidths



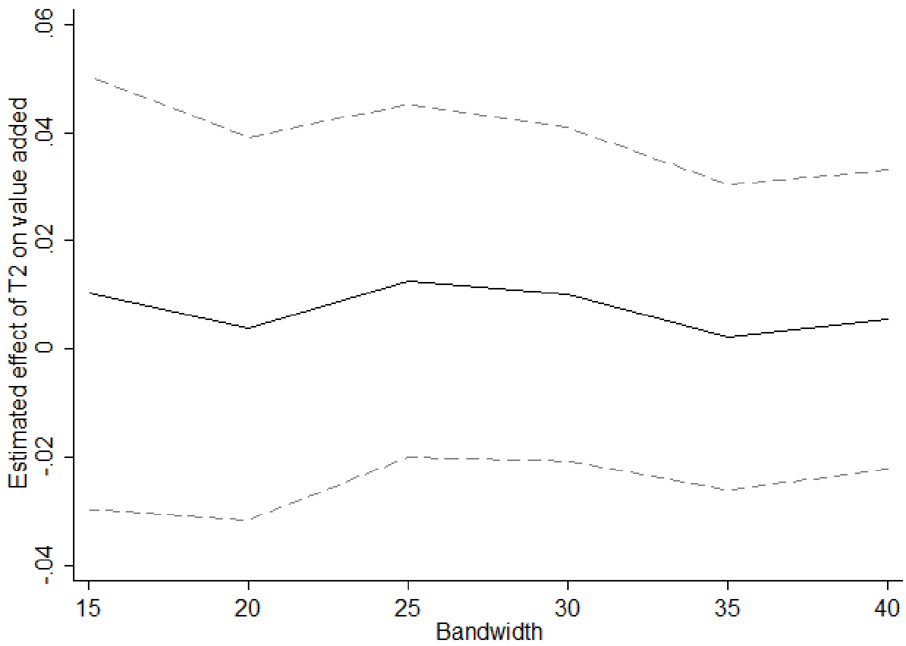
Notes: Solid lines represent estimated effects of receiving incentive 1 on school value added 2010-2011 using model 2 in the text including covariates. Dashed lines represent 95% confidence intervals.

Figure 9. Estimated effect of T2 on ENLACE score 2011 over different bandwidths



Notes: Solid lines represent estimated effects of receiving incentive 2 on ENLACE standardized score 2011 using model 2 in the text including covariates. Dashed lines represent 95% confidence intervals.

Figure 10. Estimated effect of T2 on school value-added 2010-2011 over different bandwidths



Notes: Solid lines represent estimated effects of receiving incentive 2 on school value added 2010-2011 using model 2 in the text including covariates. Dashed lines represent 95% confidence intervals.

Table 5. . Effect of receiving T1 on ENLACE test score 2011 and on school value-added 2010-2011using different subsamples

	Score 2011 (s.d.)						Value added 2010-2011					
	40	35	30	25	20	15	40	35	30	25	20	15
Primary schools												
Effect of receiving incentive for high score (T1)	-0.0226	-0.0236	-0.0310	-0.0349	-0.0323	-0.0426	-0.00896	-0.00893	-0.0269	-0.0305	-0.0260	-0.0256
	(0.0197)	(0.0209)	(0.0221)	(0.0236)	(0.0249)	(0.0284)	(0.0164)	(0.0177)	(0.0192)	(0.0212)	(0.0236)	(0.0280)
Sd. (T1)	0.0761***	0.0781***	0.0741***	0.0691***	0.0343***	0.00487	0.00138	0.0108***	0.0192*	0.0354***	0.0359***	0.0605***
	(0.0168)	(0.0172)	(0.0186)	(0.0215)	(0.0337)	(0.0185)	(0.00994)	(0.0122)	(0.0296)	(0.0215)	(0.0262)	(0.0253)
n	22253	19502	16638	13819	11049	8258	22742	19502	16638	13819	11049	8258
Primary schools (CONAFE)												
Effect of receiving incentive for high score (T1)	-0.0939	-0.104	-0.133	-0.107	-0.141	-0.0909	-0.206~	-0.167	-0.191	-0.129	-0.146	0.0341
	(0.106)	(0.113)	(0.125)	(0.134)	(0.154)	(0.180)	(0.117)	(0.126)	(0.136)	(0.147)	(0.167)	(0.195)
Sd. (T1)	0.186***	0.136**	0.155*	0.181*	0.295**	0.316*	0.0625*	0.0722*	0.109**	0.110**	0.206**	0.267**
	(0.0775)	(0.0856)	(0.126)	(0.131)	(0.130)	(0.167)	(0.0746)	(0.0813)	(0.0864)	(0.0915)	(0.111)	(0.137)
n	1246	1083	928	782	602	462	1246	1083	928	782	602	462
Primary schools (indigenous)												
Effect of receiving incentive for high score (T1)	0.00655	-0.0510	-0.0385	-0.107	-0.133	-0.0662	-0.125	-0.228*	-0.210*	-0.201*	-0.276*	-0.253*
	(0.0839)	(0.0886)	(0.0968)	(0.104)	(0.116)	(0.138)	(0.0865)	(0.0887)	(0.0926)	(0.101)	(0.112)	(0.129)
Sd. (T1)	0.0450***	0.0643**	0.0357	0.0730*	0.0538***	0.00798***	0.0287~	0.0169	0.0155***	0.0297	0.0599*	0.0649*
	(0.0247)	(0.0652)	(0.0726)	(0.0795)	(0.0289)	(0.00583)	(0.0598)	(0.0681)	(0.0100)	(0.0829)	(0.0776)	(0.0871)
n	1349	1100	918	791	647	465	1285	1031	948	791	647	495
Secondary schools												
Effect of receiving incentive for high score (T1)	0.0269	0.0357	0.0308	0.0569	0.0420	0.0676	0.0615~	0.0684~	0.0708~	0.0679	0.0481	0.0403
	(0.0360)	(0.0368)	(0.0372)	(0.0406)	(0.0444)	(0.0508)	(0.0328)	(0.0352)	(0.0373)	(0.0454)	(0.0454)	(0.0514)
Sd. (T1)	0.100***	0.0854***	0.0550***	0.0497***	0.0432***	0.0320***	0.0202***	0.0220*	0.00992***	0.131***	0.0692***	0.0294*
	(0.0411)	(0.0461)	(0.0320)	(0.0347)	(0.0393)	(0.0138)	(0.00829)	(0.0354)	(0.00460)	(0.0620)	(0.0515)	(0.0481)
n	2880	2498	2106	1741	1359	1004	2880	2498	2155	1741	1359	1004
Secondary schools (CONAFE)												
Effect of receiving incentive for high score (T1)	0.291~	0.246	0.317	0.239	0.131	0.285	-0.0291	-0.0335	0.0221	0.00197	0.0628	0.154
	(0.175)	(0.185)	(0.204)	(0.220)	(0.248)	(0.276)	(0.170)	(0.182)	(0.201)	(0.203)	(0.223)	(0.241)
Sd. (T1)	0.241~	0.260~	0.241	0.234	0.326	0.159	0.287~	0.419~	0.456~	0.395~	0.427~	0.260~
	(0.198)	(0.208)	(0.276)	(0.288)	(0.263)	(0.537)	(0.185)	(0.191)	(0.198)	(0.193)	(0.195)	(0.212)
n	221	196	162	137	108	77	221	196	162	137	108	77
Secondary schools (technical and distance learning)												
Effect of receiving incentive for high score (T1)	0.0261	0.00835	0.00934	0.00832	0.0175	-0.0283	-0.00376	-0.00528	0.0104	0.0106	0.0136	-0.004
	(0.0298)	(0.0316)	(0.0341)	(0.0364)	(0.0412)	(0.0473)	(0.0287)	(0.0306)	(0.0322)	(0.0350)	(0.0391)	(0.0431)
Sd. (T1)	0.0316***	0.0379***	0.0344***	0.0339***	0.0465***	0.0396***	0.0648***	0.0587***	0.0490***	0.0354***	0.0143***	0.0164
	(0.0171)	(0.0181)	(0.0184)	(0.0201)	(0.0228)	(0.0251)	(0.0179)	(0.0182)	(0.0183)	(0.0214)	(0.00417)	(0.0063)
n	8079	7118	5909	5040	3857	2915	8079	6927	6008	4854	3749	2915
Schools in poor municipalities												
Effect of receiving incentive for high score (T1)	-0.0350	-0.0323	-0.0245	-0.0313	-0.0455	-0.0431	-0.0520*	-0.0391	-0.0397	-0.0421	-0.0676~	-0.0485
	(0.0274)	(0.0292)	(0.0306)	(0.0327)	(0.0386)	(0.0427)	(0.0259)	(0.0275)	(0.0296)	(0.0324)	(0.0364)	(0.0422)
Sd. (T1)	0.0643***	0.0680***	0.0465***	0.0145**	0.00984***	0.0215***	0.0105*	0.0141**	0.00100	0.00813~	0.0458***	0.0474**
	(0.0251)	(0.0262)	(0.0357)	(0.0205)	(0.00225)	(0.00531)	(0.0194)	(0.0196)	(0.0206)	(0.0213)	(0.0243)	(0.0474)
n	13765	12053	10231	8523	6141	5124	13765	12053	10394	8523	6639	5124
Schools in rural areas												
Effect of receiving incentive for high score (T1)	-0.0174	-0.0203	-0.0256	-0.0394	-0.0378	-0.0280	-0.0244	-0.0134	-0.0250	-0.0244	-0.0317	-0.0129
	(0.0231)	(0.0247)	(0.0264)	(0.0282)	(0.0302)	(0.0369)	(0.0216)	(0.0230)	(0.0248)	(0.0270)	(0.0297)	(0.0350)
Sd. (T1)	0.0636***	0.0695***	0.0640***	0.0460***	0.0243	0.0101~	0.00627*	0.0161	0.0151	0.00867***	0.0284***	0.0451***
	(0.0212)	(0.0221)	(0.0253)	(0.0350)	(0.0705)	(0.0239)	(0.0154)	(0.0606)	(0.0630)	(0.00216)	(0.0194)	(0.0423)
n	19029	16637	14154	11747	9871	6598	19029	16637	14154	11747	9623	7045

Notes: Score 2011 (s.d.) is the average ENLACE score for Mathematics and Spanish obtained for each school during 2011 standardized to have mean zero and variance equal to one. Value added 2010-2011 is the change in the standardized average ENLACE scores between 2010 and 2011. Standard errors are shown in parenthesis. The table reports the point estimate and the standard error on the coefficients of receiving each of the incentives as shown in model 2 in the text using a linear model with the same slope at each side of the cutoff. It also reports the standard deviation of the coefficient obtained from the random estimations of the mixed model. Covariates included : total number of students, basic infrastructure index, school infrastructure index, educational material index, personnel index, teacher quality index, evaluation index, dropout index, internal environment index and external environment index. All models restrict the sample to schools scoring within the bandwidth specified at the top of the column and to the subsample of schools specified in each row. All models exclude schools that scored above the cutoff but not received a reward. ~ $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 6. Effect of receiving T2 on ENLACE test score 2011 and on school value-added 2010-2011using different subsamples

	Score 2011 (s.d.)						Value added 2010-2011					
	40	35	30	25	20	15	40	35	30	25	20	15
Primary schools												
Effect of receiving incentive for increase (T2)	0.0214	0.0161	0.0163	0.00851	-0.00567	0.0173	0.0254	0.0202	0.0272	0.0322~	0.0296	0.0437~
	(0.0207)	(0.0209)	(0.0227)	(0.0245)	(0.0255)	(0.0293)	(0.0166)	(0.0167)	(0.0188)	(0.0195)	(0.0223)	(0.0255)
Sd. (T2)	0.0807***	0.0680***	0.0797***	0.0812***	0.0367**	0.0204*	0.0489***	0.0336***	0.0584***	0.0444***	0.0610***	0.0626***
	(0.0196)	(0.0205)	(0.0213)	(0.0275)	(0.0407)	(0.0314)	(0.0206)	(0.0225)	(0.0217)	(0.0255)	(0.0252)	(0.0300)
n	25605	22713	19458	16027	12577	9138	25605	22713	19458	16027	12577	9138
Primary schools (comunitarian)												
Effect of receiving incentive for increase (T2)	0.0342	0.0373	0.137	0.149	0.275~	0.257	-0.0220	-0.0126	0.0968	0.0646	0.0612	-0.0221
	(0.118)	(0.127)	(0.136)	(0.146)	(0.164)	(0.190)	(0.126)	(0.137)	(0.145)	(0.158)	(0.181)	(0.211)
Sd. (T2)	0.0933**	0.0954	0.0818*	0.0677~	0.0245***	0.117~	0.148*	0.215**	0.119*	0.122	0.313*	0.420~
	(0.0840)	(0.268)	(0.0954)	(0.104)	(0.0136)	(0.130)	(0.119)	(0.125)	(0.126)	(0.166)	(0.181)	(0.190)
n	990	852	724	604	472	370	998	852	724	611	472	364
Primary schools (indigenous)												
Effect of receiving incentive for increase (T2)	-0.195*	-0.233*	-0.265*	-0.295*	-0.263*	-0.281~	0.0122	-0.0334	-0.0512	-0.0535	-0.141	-0.179
	(0.0924)	(0.0986)	(0.108)	(0.126)	(0.134)	(0.161)	(0.0882)	(0.0922)	(0.101)	(0.112)	(0.125)	(0.143)
Sd. (T2)	0.0803**	0.0648	0.123~	0.200*	0.165	0.321**	0.130***	0.118***	0.133***	0.153***	0.191***	0.159*
	(0.0733)	(0.296)	(0.154)	(0.143)	(0.236)	(0.141)	(0.0715)	(0.0753)	(0.0748)	(0.0801)	(0.0875)	(0.140)
n	1243	1042	900	730	576	410	1243	1042	900	730	576	410
Secondary schools												
Effect of receiving incentive for increase (T2)	0.0421	0.0264	0.0258	0.00977	-0.00483	-0.0115	0.0147	0.0166	0.0245	0.0482	0.0765~	0.0640
	(0.0343)	(0.0357)	(0.0372)	(0.0393)	(0.0420)	(0.0505)	(0.0277)	(0.0293)	(0.0302)	(0.0331)	(0.0408)	(0.0456)
Sd. (T2)	0.0451***	0.0404***	0.0428***	0.0479***	0.0493**	0.0629**	0.000758	0.0125***	0.0137***	0.0513***	0.153***	0.151***
	(0.0353)	(0.0376)	(0.0394)	(0.0410)	(0.0458)	(0.0619)	(0.0303)	(0.00568)	(0.00633)	(0.0366)	(0.0505)	(0.0528)
n	3977	3684	3300	2808	2246	1646	3977	3684	3300	2808	2246	1646
Secondary schools (CONAFE)												
Effect of receiving incentive for increase (T2)	0.0114	0.214	0.188	0.292	0.248	0.167	-0.218	-0.241	-0.292	-0.197	-0.263	-0.448
	(0.211)	(0.231)	(0.236)	(0.245)	(0.265)	(0.291)	(0.198)	(0.192)	(0.214)	(0.200)	(0.237)	(0.284)
Sd. (T2)	0.281~	0.466~	0.467~	0.458~	0.515	0.466	0.142~	0.0624	0.103	0.188	0.279	0.316
	(0.200)	(0.206)	(0.202)	(0.189)	(0.221)	(0.276)	(0.162)	(0.163)	(0.177)	(0.240)	(0.222)	(0.303)
n	132	111	92	79	65	50	132	111	92	79	65	50
Secondary schools (technical and distance learning)												
Effect of receiving incentive for increase (T2)	0.0137	-0.00266	-0.00421	0.0112	0.0410	0.0731	-0.0218	-0.0207	-0.0121	-0.0260	-0.0167	0.0118
	(0.0361)	(0.0379)	(0.0397)	(0.0416)	(0.0450)	(0.0495)	(0.0313)	(0.0333)	(0.0357)	(0.0368)	(0.0409)	(0.0428)
Sd. (T2)	0.141***	0.154***	0.144***	0.130***	0.101***	0.0653*	0.112***	0.130***	0.146***	0.112***	0.125***	0.0440**
	(0.0334)	(0.0335)	(0.0352)	(0.0364)	(0.0484)	(0.0854)	(0.0305)	(0.0301)	(0.0313)	(0.0350)	(0.0382)	(0.0431)
n	7750	6912	5999	5031	3976	3001	7750	6912	5999	5031	3976	3001
Schools in poor municipalities												
Effect of receiving incentive for increase (T2)	0.0128	-0.00210	-0.00520	-0.0218	-0.0177	0.00153	0.0316	0.0290	0.0321	0.0252	0.00307	0.0274
	(0.0331)	(0.0348)	(0.0363)	(0.0390)	(0.0412)	(0.0474)	(0.0286)	(0.0295)	(0.0322)	(0.0345)	(0.0390)	(0.0429)
Sd. (T2)	0.130***	0.135***	0.121***	0.116***	0.0577*	0.0827*	0.0924***	0.0810***	0.0991***	0.0971***	0.129***	0.0970***
	(0.0318)	(0.0333)	(0.0354)	(0.0426)	(0.0672)	(0.0819)	(0.0380)	(0.0388)	(0.0386)	(0.0464)	(0.0461)	(0.0422)
n	13034	11384	9683	7941	6267	4756	13034	11384	9683	7941	6267	4756
Schools in rural areas												
Effect of receiving incentive for increase (T2)	0.0137	0.00379	0.00417	-0.00825	-0.0206	-0.0121	0.0137	0.00379	0.00417	-0.00825	-0.0206	-0.0121
	(0.0269)	(0.0282)	(0.0301)	(0.0329)	(0.0354)	(0.0400)	(0.0269)	(0.0282)	(0.0301)	(0.0329)	(0.0354)	(0.0400)
Sd. (T2)	0.102***	0.0945***	0.0944***	0.100***	0.0459*	0.0314	0.102***	0.0945***	0.0944***	0.100***	0.0459*	0.0314
	(0.0255)	(0.0281)	(0.0293)	(0.0336)	(0.0585)	(0.104)	(0.0255)	(0.0281)	(0.0293)	(0.0336)	(0.0585)	(0.104)
n	17885	15471	13122	10763	8459	6475	17885	15471	13122	10763	8459	6475

Notes: Score 2011 (s.d.) is the average ENLACE score for Mathematics and Spanish obtained for each school during 2011 standardized to have mean zero and variance equal to one. Value added 2010-2011 is the change in the standardized average ENLACE scores between 2010 and 2011. Standard errors are shown in parenthesis. The table reports the point estimate and the standard error on the coefficients of receiving each of the incentives as shown in model 2 in the text using a linear model with the same slope at each side of the cutoff. It also reports the standard deviation of the coefficient obtained from the random estimations of the mixed model. Covariates included: total number of students, basic infrastructure index, school infrastructure index, educational material index, personnel index, teacher quality index, evaluation index, dropout index, internal environment index and external environment index. All models restrict the sample to schools scoring within the bandwidth specified at the top of the column and to the subsample of schools specified in each row. All models exclude schools that scored above the cutoff but not received a reward. ~ $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 7. Effect of receiving T1 AND T2 on ENLACE test score 2011 and on school value-added 2010-2011

	Score 2011 (s.d.)						Value added 2010-2011					
	40	**35**	**30**	**25**	**20**	**15**	**40**	**35**	**30**	**25**	**20**	**15**
Response surface regression discontinuity												
Effect of receiving incentive for high score (T1)	0.00208	0.0107	0.0146	0.0291	0.0134	0.0103	0.00942	0.00971	0.00446	0.0208	0.0150	0.0176
	(0.0168)	(0.0189)	(0.0218)	(0.0249)	(0.0292)	(0.0427)	(0.0163)	(0.0185)	(0.0197)	(0.0240)	(0.0251)	(0.0377)
Sd. (T1)	0.132***	0.137***	0.173***	0.196***	0.138***	0.213***	0.0798***	0.0981***	0.0585***	0.108***	0.0455~	0.0188
	(0.0164)	(0.0188)	(0.0240)	(0.0298)	(0.0380)	(0.0536)	(0.0233)	(0.0238)	(0.0363)	(0.0325)	(0.0761)	-
Effect of receiving incentive for increase (T2)	-0.00775	-0.00852	-0.00298	-0.0461~	-0.0557~	-0.0801~	0.0169	0.0191	0.0186	-0.0154	-0.0207	-0.0424
	(0.0185)	(0.0204)	(0.0239)	(0.0280)	(0.0317)	(0.0459)	(0.0180)	(0.0198)	(0.0238)	(0.0267)	(0.0298)	(0.0434)
Sd. (T2)	0.0830***	0.0965***	0.109***	0.105***	0.0652***	0.106***	0.135***	0.140***	0.198***	0.195***	0.220***	0.294
	(0.0160)	(0.0181)	(0.0238)	(0.0292)	(0.0503)	(0.0694)	(0.0202)	(0.0218)	(0.0250)	(0.0286)	(0.0324)	-
n	21486	16878	12518	8714	5699	2926	21486	16878	12518	8714	5699	2926
Frontier regression discontinuity Ra|b (T2=1)												
Effect of receiving two incentives (T1 & T2)	-0.0199	-0.0146	-0.0394	-0.0627	-0.0571	-0.0709	0.00699	0.0185	-0.00660	-0.0274	-0.0126	-0.0219
	(0.0424)	(0.0456)	(0.0486)	(0.0528)	(0.0594)	(0.0707)	(0.0414)	(0.0441)	(0.0476)	(0.0522)	(0.0582)	(0.0679)
Sd. (T1 & T2)	0.141***	0.146***	0.125***	0.130***	0.181***	0.265***	0.109***	0.107***	0.109***	0.131***	0.148***	0.176***
	(0.0418)	(0.0417)	(0.0479)	(0.0557)	(0.0546)	(0.0599)	(0.0459)	(0.0478)	(0.0509)	(0.0515)	(0.0549)	(0.0634)
n	5558	4883	4220	3531	2839	2158	5558	4883	4220	3531	2839	2158
Frontier regression discontinuity Ra|d (T1=1)												
Effect of receiving two incentives (T1 & T2)	-0.0364	-0.0542	-0.0411	-0.0788~	-0.117*	-0.135*	0.0242	0.00638	0.0484	0.0423	0.0178	0.0194
	(0.0375)	(0.0390)	(0.0420)	(0.0456)	(0.0498)	(0.0546)	(0.0314)	(0.0336)	(0.0365)	(0.0396)	(0.0442)	(0.0515)
Sd. (T1 & T2)	0.249***	0.250***	0.277***	0.286***	0.253***	0.103*	0.128***	0.157***	0.153***	0.143***	0.149***	0.170***
	(0.0333)	(0.0342)	(0.0371)	(0.0390)	(0.0503)	(0.0960)	(0.0283)	(0.0294)	(0.0330)	(0.0366)	(0.0407)	(0.0449)
n	6024	5378	4659	3889	3093	2263	6024	5378	4659	3889	3093	2263

Notes: Score 2011 (s.d.) is the average ENLACE score for Mathematics and Spanish obtained for each school during 2011 standardized to have mean zero and variance equal to one. Value added 2010-2011 is the change in the standardized average ENLACE scores between 2010 and 2011. Standard errors are shown in parenthesis. Response surface regression discontinuity estimates are the point estimate and the standard error on the coefficients of receiving each of the incentives as shown in model 3 in the text using a linear model with the same slope at each side of the cutoff. It also reports the standard deviation of the coefficient obtained from the random estimations of the mixed model. Frontier regression discontinuity estimates are the point estimate and the standard error on the coefficients of receiving each of the incentives as shown in model 4 and 5 in the text using a linear model with the same slope at each side of the cutoff. It also reports the standard deviation of the coefficient obtained from the random estimations of the mixed model. All models include the following covariates: total number of students, basic infrastructure index, school infrastructure index, educational material index, personnel index, teacher quality index, evaluation index, dropout index, internal environment index and external environment index. All models restrict the sample to schools scoring within the bandwidth specified at the top of the column. All models exclude schools that scored above the cutoff but not received a reward. ~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001