European University Institute

MAX WEBER PROGRAMME FOR POSTDOCTORAL STUDIES

EUI WORKING PAPERS

MWP 2015/14
Max Weber Programme

Rules Rather than Discretion: Teacher Hiring and Rent Extraction

Ricardo Estrada

European University Institute
Max Weber Programme

Rules Rather than Discretion: Teacher Hiring and Rent Extraction

Ricardo Estrada

EUI Working Paper **MWP** 2015/14

ISSN 1830-7728

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract
Because of data limitations, there is little empirical research on how firms conduct hiring and the merits of different recruitment strategies. In this paper, I take advantage of a unique setting that allows me to compare the quality (value-added to student achievement) of the teachers hired in a discretionary process led by the teachers' union in Mexico with those hired on the basis of a screening rule. My results show that the teachers' union selects applicants of a considerably lower quality than those selected using a standardized test, despite the fact that the test has no power to predict teacher quality. I find evidence that the results are not explained by the self-selection of high-quality teachers to follow the test-based process. The combination of these results indicates that the teachers selected through the discretionary process are from the bottom of the distribution of applicant quality. My analysis also reveals that joint committees of state officials and union representatives allocate teachers hired in this way to schools in more "desirable" localities, but with similar pre-treatment trends in outcomes. Findings are consistent with standard models of rent extraction.

Keywords
Hiring methods, teachers' unions, school quality, teacher hiring, rent extraction.

JEL codes: I21, J51, M51.

I thank Marc Gurgand and Jérémie Gignoux for advice and encouragement. I am indebted for insightful comments to Yann Algan, Joshua Angrist, François Bourguignon, Alejandro del Valle, Pascaline Dupas, Francisco Ferreira, Paul Glewwe, Andrea Ichino, Fabian Lange, and participants at seminars at the European University Institute, Paris School of Economics and Colegio de Mexico, the conferences NEUDC Harvard, LACEA Colegio de Mexico and MIEDC Minnesota, and workshops at Lisbon ISEG, LACEA-IEN Grade, ZEW Mannheim University, Oxford Blavatnik and LACEA-PEG Universidad de los Andes. I am also grateful to Jorge Santibanez for his insights about the Mexican education system and Kenneth Houngbedji for sharing Stata routines. I thank the staff at Mexicanos Primero for kindly providing me with access to data. Financial support from CONACYT and the Max Weber Programme at EUI is gratefully acknowledged.

Ricardo Estrada
Max Weber Fellow, 2014-2015
European University Institute, Via dei Roccettini, 9, 1-50014 San Domenico di Fiesole,
ricardo.estrada@eui.eu.

1 Introduction

How should organizations hire workers? The main challenge for employers is that they typically have imperfect information about applicant quality. In the absence of an objective measure that captures quality, organizations may rely on the subjective judgement of staff members (or departments) with informational advantages to evaluate applicants. However, giving discretion to individuals to take hiring decisions may open the door to favoritism. In such cases, employers face a trade-off between recruiting using rules that are second-best predictors of quality but hard to manipulate, or relying on the choices of an agent that can potentially abuse discretion. Our understanding of such a trade-off in empirical settings is limited though, because conducting a proper analysis requires both exogenous variation in recruitment methods –in otherwise similar organizations– and the possibility of measuring worker quality.

In this paper, I take advantage of a unique setting that allows me to compare the quality (value-added to student achievement) of the teachers hired in a discretionary process led by the teachers' union in Mexico to those hired on the basis of a screening rule. Specifically, I estimate the causal effect on student achievement of allocating to a school a new teacher selected by the teachers' union in a discretionary process versus one selected through a standardized test with no predictive power for teacher quality. For the empirical analysis, I exploit: the variation in hiring introduced by a recent reform; a new personnel data set that allows me to identify new teachers by hiring status and that I link to a panel of student scores in a national standardized exam; and, for causal inference, a centralized allocation process of teachers to schools that does not depend on past trends in school outcomes.

Teachers are an interesting case to learn about hiring with imperfect information. Ample research shows both that variance of teacher quality is high and that identifying at the moment of hiring who actually is (could be) a good teacher is difficult (Hanushek and Rivkin, 2006). The problem for policy makers (and parents) is that the characteristics that can typically be observed when recruiting are, at best, modest predictors of teacher quality (Rivkin, Hanushek and Kain, 2005; Rockoff, Jacob, Kane and Staiger, 2011). The participation of the teachers' union in the selection of new teachers may ease this informational problem if current teachers have superior information or a higher ability to identify teacher quality using broader criteria.[1] However, the use of discretion also opens the door to potential favoritism: indeed, the variation in hiring that I use in this paper is the result of a reform introduced following concerns over the abuse of discretion in the prevalent hiring regime.

In September 2013, the Mexican Senate passed a constitutional amendment that makes the hiring of new teachers for public schools on the basis of standardized testing mandatory. "The inheritance and sale of jobs has ended," Education Secretary Emilio Chuayffet said. "Merit is the ideal means of access

[1] In support of this idea, Jacob and Lefgren (2008) find that school principals do well at identifying teachers in the extremes of the quality distribution. Also, Rockoff and Speroni (2011) use NYC public schools data to show that subjective evaluations of new teachers predict the achievement of future students (controlling for student characteristics and past achievement).

to, and progress in, a teaching career." (Weissenstein, 2013). This piece of legislation was meant to end an era in which the teachers' union had a leading voice in teacher hiring. One contribution of the paper is to evaluate this ambitious reform in a large education system (with around 1 million teachers).

In Mexico, state governments operate the public primary (grades 1 to 6) and junior-secondary (grades 7 to 9) schools, while the federal government sets the national curricula and provides states with the bulk of funding. Teacher hiring is centralized at the state level and is not carried out independently by schools.

The 2013 reform did not introduce teacher testing to the public education system for the first time. What it did was to scale-up a test-based recruitment process adopted nationwide after a previous reform that took place in 2008, which means that from to 2008 to 2013 both hiring systems co-existed in Mexico.[2] In this paper, I focus on a set of junior high schools that received new teachers in the academic year 2010-2011 (henceforth referred to as 2010 for simplicity's sake). Although the test-based hiring was introduced in 2008, not all schools immediately received a test-hired teacher. Both because not all new teachers were selected using the test-examination and because not all schools have vacant positions every year. So, I compare schools that received either only test-hired or only discretionary-hired teachers in 2010 and had not received a test-hired teacher before.

I am able to identify the 2010 cohort of teachers recruited through each of the hiring systems by using extensive data of school personnel compiled following a recent mandate of the Mexican Federal Congress. I match the teachers' data to panel data (which I compiled) on school characteristics and student scores in a national standardized test from 2005 to 2010 (five years before and one year after the allocation of teachers). I focus on Telesecundaria schools, a public system of junior secondary education (grades 7 to 9) which accounts for 20% of total enrollment at this schooling level. Telesecundarias are small schools, catering to small communities, with one teacher per classroom.

To identify a causal effect, I rely on an allocation process of teachers to schools that I show, in line with anecdotal evidence, does not depend on past school trends in school outcomes. Figure 1 shows the raw variation in the data (the evolution of student achievement in schools that received new teachers in 2010) and announces my main results. Before treatment, schools that received test-hired teachers had on average lower performing students than schools that received discretionary-hired teachers. But, crucially for the identification of a causal effect in a difference-in-difference model (based on the parallel-trend assumption), both sets of schools follow similar trends in outcomes during the five-year period before treatment. After treatment, the outcomes converge.

My difference-in-differences estimates confirm that the allocation of test teachers has a positive and sizable effect on student achievement. Moving from no test teachers in a school to only test teachers increases the school's mathematics test score by .52 SD and the Spanish score by .31 SD, a result that is

[2]The name of the test-based recruitment is (in Spanish) Concurso Nacional de Asignacion de Plazas Docentes. The states of Michoacan and Oaxaca, strongholds of CNTE a smaller teachers' union, did not adopted the test-based hiring.

statistically significant at the five-percent level in both cases. Furthermore, going from not having test teachers in the school to having only test teachers reduces by 5.8 percentage points the share of students detected copying in exams.[3] Results are robust to different specifications and checks –which I describe in detail in Section 5.4.

I then provide evidence on whether the large quality gap between test and discretionary teachers is due to the high screening efficiency of the test method and/or the self-selection of high-quality applicants to this method. With respect to the first channel, there is a substantial amount of literature that shows that the relation between teacher testing and teacher quality is inconsistent; with positive, insignificant and negative results (see for reviews: Hanushek and Rivkin 2006, Wayne and Youngs 2003 and Glewwe, Hanushek, Humpage and Ravina 2011).[4] Even when significant, teacher test scores are modest predictors of teacher quality. A group of high-quality studies that find positive (significant) results estimate that a 1 SD increase in teacher test scores leads to around .04 SD increase in mathematics –with a lower effect for Reading (Clotfelter, Ladd and Vigdor, 2006; Goldhaber, 2007; Rockoff, Jacob, Kane and Staiger, 2011).

To investigate the predictive power over teacher quality of the specific test used to select teachers in the test examination, I estimate a value-added model in which I regress student achievement on teacher test scores while I control for past levels of achievement in the school using a school fixed-effects model. My findings suggest that the exam used in the test-based selection has no predictive power over teacher effectiveness (the estimated coefficients of interest are both small and insignificant). The estimation of this model presents some limitations due to a small sample size and truncated data on the main independent variable (I do not observe the school outcomes of the test applicants not hired). Reassuringly, the estimated coefficients are in line with those reported in high-quality studies (though are not statistically significant). Even if significant, the magnitudes of the point estimates are too small in comparison with the quality gap found between test and discretionary teachers.

I also find evidence that the observed quality gap between test and discretionary teachers is not completely explained by the self-selection of high-quality teachers to the test-based process. To do so, I repeat the main estimation but restricting it now to test teachers and discretionary teachers who also applied for teaching positions through the test selection process. Focusing on this similar pool of applicants (in the sense that both groups decided to self-select to test application), I observe the same results as before.

Summing up, the findings here presented show that the discretionary-hired teachers do considerably worse than the test-hired teachers, even though the standardized exam used in the test hiring bears no relation to teacher quality. Furthermore, when I compare the efficacy of the two methods on a similar pool

[3]The Federal Secretary of Education measures exam cheating using a detection algorithm designed to give lower-bound estimates for direct copying, e.g. one student copying from another or a larger group of students (and potentially teachers) exchanging responses during the application of the exam.

[4]Tests focused on subject knowledge are better predictors of teacher quality (Glewwe, Hanushek, Humpage and Ravina, 2011).

of applicants, those who self-select for the test application, the discretionary method is also found to lead to the hiring of lower quality teachers. The combination of these results indicates that the discretionary teachers are selected from the bottom of the distribution of applicant quality. Although striking, this pattern is consistent with the main motivation of the reform: avoiding the inheritance and sale of teaching positions.

A basic principal-agent framework can be useful to interpret the results. Note that a hiring agent that maximizes private rents –rather than worker quality– has the incentive to hire low-quality applicants. If the agent decides to put a teaching position up for sale (demands a bribe of applicants), the side payment is maximized by selecting the worst applicant, the one who has the lowest outside option in the labor market and hence the incentive to pay the highest bribe. Similarly, for a retiring teacher who maximizes family income, the optimal solution is to select the relative with the worst outside option.

Furthermore, I find that joint committees of state officials and union representatives allocate discretionary teachers to schools located in more "desirable" localities; i.e. closer to the state capital, less poor and with a higher coverage of public services –but with similar pre-treatment trends in outcomes. A regression analysis reveals that what matters to predict the allocation of discretionary over test teachers to a schools are the characteristics of the locality, but not school outcomes (both in levels and trends).

Oyer and Schaefer (2011) assess the literature on personnel economics and argue for research directed at improving our understanding of how firms conduct hiring and the merits of different recruitment strategies. This paper aims to contribute to fill this gap. I use a unique empirical setting to compare the quality of workers selected using two different recruitment methods and illustrate the structural tension that may arise when an agent screens workers in an economy with asymmetric information about worker quality.

In related studies, Durante, Labartino and Perotti (2014) find that the decentralization of hiring decisions in Italian universities increases familism in areas with a low civic capital and, regarding promotions, Zinovyeva and Bagues (2015) find that (external) evaluators are more likely to promote candidates with whom they have ties in the Spanish university system (strong links lead to the selection of lower-quality candidates, while weak links to higher-quality candidates). Duflo, Dupas and Kremer (2015) observe that training parents in school governance in Kenya decreases the probability that school principals hire a teacher's relative.

I also contribute to the literature on the effects of teachers' unions on school quality, a topic in which there is a vivid controversy on whether and how teachers' unions push to set school inputs and policies that maximize school quality or that deviate resources from this objective. This literature is relatively thin though, as there are important empirical challenges to tackle these questions. First is the problem of identification or the lack of appropriate variation to separate the causal effects of unions from other school characteristics that affect both the likelihood of unionization and student achievement.

Second is the absence of detailed data on specific mechanisms through which unions shape the education production function. In a seminal paper, Hoxby (1996) uses (time and spacial) variation in unionization among school districts in the U.S. and finds teacher unionization is associated with an increase in schools' budget, teacher salaries, student-teacher ratios, and drop-out rates. Lovenheim (2009) uses an improved measure of unionization in school districts in three Midwestern states and finds no effect of unions on teacher pay, per-student expenditures or dropout rates. Kingdon and Teal (2010) exploit within-pupil (across-subjects) variation in union status and finds that unionization is associated with a decrease in student achievement in private schools in India. In this paper, I evaluate how the teachers' union affects school quality through a specific mechanism: teacher hiring.

2 Teacher Hiring

2.1 Conceptual Framework

Before presenting the institutional context, I outline a very simple principal-agent model to illustrate the structural tension that may arise when an agent screens workers in an economy with asymmetric information about worker productivity.

Consider a principal who engages an agent to hire a teacher from a continuum of applicants with quality (q_i) uniformly distributed on a segment represented by $[0, 1]$ and expressed in (normalized) monetary terms. Suppose the principal does not observe applicant quality, while the agent might or might not. Applicants know their own quality (outside option in the labor market). The process of filling the vacancy is straightforward: 1) the principal posts a vacancy with a positive (and exogenous) wage w, 2) would-be teachers with an outside option lower (or equal) to the posted wage apply for the vacancy ($q_i \leq w$) and 3) the agent selects applicant i of quality q_i from the segment $[0, w]$.

Assume the principal cares only about teacher quality, while the agent may derive positive utility from teacher quality or from a rent (r_i) that the agent can extract if the posted wage is greater than the hiree's quality (say $r_i = w - q_i$). Consider first the case in which the agent maximizes teacher quality. Then, if the agent observes quality (is an efficiency-enhancer), she selects the highest-quality applicant ($q_i^E = w$), while if she does not observe quality (is a naive hirer), her choice is, on expectation, the mean (quality) applicant ($\mathrm{E}\left[q_i^N\right] = \frac{w}{2}$]. In contrast, if the agent maximizes private rents (is a rent-seeker), then her optimal solution is to select the worst-quality applicant ($q_i^R = 0$). For intuition, suppose the agent decides to put a teaching position for sale (demands a bribe of applicants). In this case, the worst-quality applicant, the one who has the lowest outside option in the labor market, is the one who has the incentive to pay the highest bribe. Similarly, for a retiring teacher who maximizes family income the optimal solution is to push for the selection of her relative with the worst outside option in the labor market.[5]

[5] In a related framework, Hoxby (1996) outlines that teachers demand unions to influence the educational production function either because 1) in an efficiency-enhancing model of participation, teachers share the objective function of adminis-

In this simple setting, the agency problem depends on the agent's type: efficiency-enhancer, naive or rent-seeker, with clear empirical implications for each case. Note that if the principal knows that the agent is a rent-seeker, she is better off replacing the agent for a very basic selection rule: picking an applicant at random.

One may wonder to what degree the agency problem depends on the static nature of the outlined model. The capacity of the agent to extract a rent may be constrained if the game is repeated and the principal can observe ex-post, at least partially, teacher quality. A simplistic way to bring a dynamic flavor into this setting is to suppose that the agent cares about both teacher quality and rents, and –with Cobb-Douglas preferences– to define a parameter α (with values strictly between 0 and 1) that captures both the agent's taste for teacher quality and the internalization of any future cost that the principal may impose on the agent for hiring low-quality teachers. So, the better the principal's technology for ex-post identification of teacher quality, the higher would be α. The same goes for the agent's capacity to effectively punish the agent for selecting low-quality candidates or to design incentives that reward the hiring of high-quality applicants. This implies that the described agency problem will be more relevant in environments with weak institutions.

2.2 Teachers' union: SNTE

State governments (31) operate the public primary (grades 1 to 6) and junior-secondary (grades 7 to 9) schools in Mexico, while the federal government sets the national curricula and provides the bulk of funding.[6] Public schools account for around 90% of total enrollment in primary and junior-secondary education. Teacher hiring and allocation to schools are centralized at the state level and are a responsibility of the state Ministries of Education. In this task, though, the teachers' union plays a key role.

The National Educational Workers Union (SNTE by its Spanish acronym) is a national organization, formed in 1943, with more than one million members. Although there are other small teachers' unions in some states, SNTE affiliates almost all teachers, principals and workers in primary and junior-secondary public schools, public teachers' schools, and the state and federal Ministries of Education. SNTE is organized in a national executive committee and 55 regional sections (around two sections per state).

Both affiliation and payment of fees (around 1% of base salary) to the union is mandatory and automatic for all teachers in public elementary schools, inclusive of new teachers. There is little accountability on the way that the SNTE leadership disburses the collected fees (Santibanez and Jarillo, 2007). Though in February of 2013 the national leader of the union (Elba Esther Gordillo) was arrested on embezzlement charges (of around 200 million dollars).

trators and parents (maximization of student achievement), but have better information about input efficiency or internalize externalities than the others neglect; or alternatively 2) in a rent-seeking model, teachers have a different objective function than administrators and parents, and hence a desire to set the school inputs that maximize their own objectives.

[6]The federal government is also in charge of managing public schools in the Federal District, where the capital of the country is located.

The teachers' union has an important say in the operation of the public education system. For example, union leaders are formally represented in the state committees that (jointly with education officials) define the hiring, allocation and promotion of teachers in schools. Through these joint committees, the union also influences the appointment of school principals and regional supervisors.

More broadly, the union is actively involved in the political system. SNTE has traditionally supported the PRI, but formed alliances with the PAN when this right-wing party arrived at the presidency of Mexico in 2000, and in 2005 created its own party, named PANAL. As a result, union leaders are regularly elected as representatives and senators, and hold leadership positions in the education committees of both the Chamber of Representatives and the Senate (Santibanez, 2008).

The participation of union leaders is not limited to the legislative power. Santibanez (2008) argues that the union leadership trades electoral and political support for the appointment of union members to middle and high-ranking positions in the federal and state ministries of education. More visibly, the head of the Federal Under-Secretary for Basic Education in the 2006-2012 administration was the in-law of the unions national leader of the time

The political clout of SNTE reflects in the government budget for education and the labor conditions of teachers. Public expenditures in education amounts to 5.4% of GDP and 20% of total public expenditures –though still short of the minimum 8% of GDP mandated by an SNTE-backed amendment to the Federal Education Law approved in 2002. Tellingly, 86% of the education budget for primary schools is spent on teachers' salaries (the highest proportion among OECD countries) (OECD, 2014). Such a level of expenditure translates into high relative wages for public school teachers.

2.3 Teacher Compensation

Table 1 compares the (mean) wages of primary and junior secondary teachers in public and private schools to those of the other college-educated workers in the labor market (using information from the National Labor Force Survey, ENOE).

First, it stands out that public school teachers have on average a higher wage than private school teachers (both monthly and hourly). The raw difference in the mean (hourly) wage is around 25% (and 21% after controlling for age and gender).

Second, though teachers from primary and junior-secondary public schools tend to have a lower monthly wage than the rest of college-educated workers, the scenario is different once the number of hours worked is taken into account. The mean hourly wage of the elementary school teachers is around 36% greater than the mean wage of the other college-educated workers (31% after controlling for age and gender), and it is similar to that of the teachers from other schooling levels and workers occupied as managers. The better pay of public school teachers in Mexico contrasts to the norm in the rest of Latin America (Mizala and Ñopo, 2014).

In addition, public school teachers have better non-wage benefits than private sector workers, including the entitlement to longer paid vacation time, a higher end-of-the-year bonus and practically guaranteed labor security (dismissal cases are rare, public school teachers facing conflict with school administrators or parents are typically transferred to an another school).

The presented statistics do not consider selection issues, but indicate that teachers are relatively well remunerated and that a career as a teacher in the public service offers, at least for some sectors of the population, better working conditions than those in the alternatives.

2.4 Discretionary Hiring

State Ministries of Education are responsible for the hiring of new teachers for public schools. By new teachers, I mean those entering the teaching profession in the public education system. This definition excludes incumbent teachers who transfer from one public school to another. The hiring of new teachers is done at the teacher-type level (primary school teacher, mathematics high school teacher, etc.) and in principle is not related to specific vacancies at schools. Though the characteristics of the hiring process vary by state, there are important patterns in common, including the participation of the teachers' union.

The teachers' union plays a prominent role in the hiring process by directly selecting a share of the total number of new teachers in every state (Guevara and Gonzalez, 2004). Acting as a de facto hiring agent is actually a common function among unions in the public sector in Mexico. The federal law that regulates the labor relations for public employees establishes that unions have the right to directly select 50% of the hirees for both new (permanent) positions and vacant positions that cannot be filled by a current employee.[7]

There is limited formal information about how exactly the teachers' union runs this selection process. The union has ample discretion over who selects for hiring, though it must comply with schooling requirements in terms of college education and type of degree.[8]

A well established practice is the entitlement of retiring teachers to select a direct relative (offspring) for hiring. A central argument used by local union leaders in their public opposition to the 2008 reform was that the examination would remove the union members' right to pass on their position (Elizondo, 2011).

Strong and widespread media outlets and policy reports denounced the selling of teaching positions. A national survey among elementary teachers found that one-third of interviewees thought that selling of teaching positions was a frequent practice, while another third said it was done with limited frequency

[7] Prior to the 2008 reform, it was also considered that the teachers' union would informally control in a majority of states, through its influence in the appointment of state officials, the share of hires done directly by the state ministries of education (see, for example, Santibanez (2008)). In support of these criticisms, a 2003 report carried out by the Federal Ministry of Education found that 19 of 32 states would not have in place any formal mechanism to select new hirees and 5 would only conduct formal evaluations of applicants in a restricted manner (Guevara and Gonzalez, 2004).

[8] New hirees must have university-level studies, though not necessarily a degree at the moment of hiring. Hirees must be either graduates from teachers schools or from fields of studies related to the subject of the teaching position. No teaching certificate is required for hiring.

(Este Pais, 2005). An internal document of the teachers' union in the state of Oaxaca recently leaked to a national newspaper estimated the selling price of a teaching position in a Telesecundaria school –like those analyzed in this paper– to be up to 300,000 pesos (around 19,000 USD) (Del Valle, 2015).

2.5 Test-based Hiring

In 2008, the Federal government, under the umbrella of a broader agreement with the teachers' union, introduced a plan to open to competitive examination all vacant teaching positions in public primary and junior secondary education in the country. I here explain the mechanics of the new examination.

Competition is open to candidates willing to enter the teaching profession in public schools and current teachers with temporary or part-time contracts. There are hiring quotas for each group. In this paper, I focus only on the recruitment of new teachers. Hiring is based on a national, standardized test held before the beginning of the academic year. There is one exam for each type of teaching position (e.g. primary school teacher, mathematics junior high school teacher, etc.). The standardized exam is designed to measure cognitive skills, knowledge of the teaching subject, mastery of teaching methods and ethics.

Candidates are ranked by state and teacher type according to their exam results or, if states opt for it, a weighted average of the test score and other criteria (often undergraduate GPA). The number and type of available teaching positions by state and the exam results are widely publicized by media outlets and are available on a dedicated web page (concursonacionalalianza.sep.gob.mx). Civil society organizations participate as monitors in different stages of the process, more visibly in the exam application. The teaching positions open to competition are not associated with specific schools. Some types of teaching positions are restricted to graduates of teacher training schools or from specific college majors.

The reform met with strong opposition from state officials and local union leaders. Compromising, only new payroll positions funded by the Federal Government were filled through the test-based recruitment initially, though it was expected that progressively more vacancies were opened to test-based hiring.[9] Almost all states began to use test-based hiring to fill some vacancies in 2008 (30 of 32, including the Federal District), until the 2013 constitutional reform made testing the mandatory mechanism to fill all vacancies starting in 2014. According to figures from the Federal Secretary of Education, from the 22,546 full-time vacancies opened to test hiring in 2010, the cohort that I study, 34% corresponded to new positions and the rest to existing payroll positions. There is no public information about the total number of new teachers hired through discretionary recruitment. They could amount to around 85% of all teachers hired in 2010, according to my estimates.[10]

[9] The teachers' union agreed to cede its selection entitlement over the 50% of the federally-funded new payroll positions.
[10] In the Telesecundaria System.

2.6 Telesecundaria Schools

I model the link between teachers and students at the school level for identification and data restrictions. Hence, I focus my empirical analysis on Telesecundaria schools, a system of public junior secondary education (Grades 7 to 9). Telesecundaria are small schools catered to small communities. The typical school in my sample has 83 students, 4 classrooms and is located in a locality with 1,000 inhabitants – all median values. The small school size should increase the likelihood that I find a statistical significant teacher effect at the school level.

The Telesecundaria system was created originally to serve localities with under 1,500 inhabitants, but over time expanded to larger rural communities and suburban areas. According to figures from the Federal Secretary of Education, around 1.26 million students attended 18,000 Telesecundaria schools in 2010, which amounts to 20.6% of total enrollment in junior secondary education.

Being mostly rural, Telesecundaria students tend to face more disadvantaged conditions than the average junior high school student. For example, in 2010 the average poverty rate in the localities where the Telesecundarias in my sample are located was 62%, while the national poverty rate was 46%,according to the National Council for the Evaluation of Social Policy (CONEVAL).

Telesecundarias have one teacher per classroom, in contrast to general high schools which have one teacher per topic. Instead of specialist teachers, Telesecundarias rely heavily on IT teaching support. The television programs that the Federal Secretary of Education produces specifically for this school system fill approximately 2 of the 6 hours of the school day. Hence, the effect of teacher quality in Telesecundarias is likely lower than in educational systems in which teachers play a larger role in the classroom.

3 Data

3.1 Enlace Exam

My main outcomes come from a national standardized test (Enlace) that students take at the end of the academic year. I use the individual test results to construct a panel dataset of school scores from 2005, the first year that the exam was given, to 2010 (five years before and one year after the treatment of interest).[11]

I observe 9th-grade scores for the whole period and 7th and 8th-grade scores from the school year 2008, when students from these grades started to sit at the exam.[12] The test measures learning in mathematics, Spanish and a rotating subject, which I exclude from the analysis. Enlace scores are standardized at the national level with mean 500 and standard deviation 100. I use a panel of 9th-grade scores in my main

[11] For simplicity, I will refer to the school year 2005-2006 as 2005 and so on, though the Enlace results from the 2005 school year correspond to the test given in the second quarter of the 2006 calendar year.

[12] The 9th-grade exam assessed materials of grades 7th to 9th before 2007, while after this year focuses in 9th-grade materials.

estimations because of the larger time dimension (2005-2010). However, I also present results using the data as a panel of classrooms for the 2008-2010 period.

The Mexican Evaluation of Scholastic Achievement of Educational Institutions (Enlace) is designed to assess the overall educational system and, hence, there is no bearing for students on GPA or graduation. However, Enlace results are widely reported by media outlets and non-governmental organizations.

Also, since 2009, the Federal Secretary of Education delivers monetary bonuses to teachers of high-performing classrooms and schools within specific categories. Bonuses are distributed to teachers (and school principals) of classrooms or schools in the top 15% of the score distribution, and of classrooms in the top 15% of the score gains distribution. Schools are classified by state into categories defined by locality characteristics (urban/rural and with high/low marginalization) and school type (general/technical/telesecundaria/etc.). A teacher can receive a bonus ranging from $2,000 up to $20,000 pesos. This is around USD PPP 260 and 2,600, respectively; or 16% to 160% of the mean monthly wage of the new teachers in my sample.

The publicity and the bonuses provide school agents with incentives to perform better and makes Enlace a medium-stake test.

3.1.1 Detection of Cheating

Students take the Enlace exam in two school days towards the end of the academic year. Each State Secretary of Education allocates one exam coordinator per school to oversee the implementation of the test alongside the school principal. The school principal selects one teacher per classroom to monitor the students during the exam. It is forbidden for teachers to monitor the classes they teach.[13] At the end of each day, the monitoring teachers must turn in the response sheets to the exam coordinator and the school principal, who pack the answer sheets into sealed boxes at the end of the second and final day of the exam. Information sheets distributed to principals and teachers state the subsequent use of a computer software to detect copying among students and provision of exam responses by a third-party.

The Federal Secretary of Education runs software to detect test cheating using two statistical tools commonly used for this purpose, the K-Index and the Error Similarity Analysis (ESA) Index.[14] Both methods measure unusual agreement between the incorrect answers of two examinees in a multiple-choice test and, as both are based on a binomial distribution, have a fairly similar general structure. The focus on common incorrect answers comes from the idea that the number of similar correct answers increases with students' true achievement level, while the identical selection of responses given as distractors is informative of copying.

[13] School principals should guarantee that at least two parents per classroom attend the exam as external observers. I do not have information about how extensively this policy is implemented.

[14] The Educational Teaching Service (ETS) routinely uses the K-Index to detect cheating in the several examinations they perform (SAT, GRE, GMAT, etc.), while the ESA Index is the basis for the, commercially available, *Scrutiny!* software. A detailed description of the methods can be found at Holland (1996).

The two indexes are designed to give lower-bound estimates for a specific form of cheating: direct copying, e.g. one student copying from another or a larger group of students (and potentially teachers) exchanging responses during the exam. Even in this case, copying will go undetected if it is restricted to a few answers (relatively to the total number of wrong responses) or if the source of copying does not have incorrect responses. Moreover, both methods are unlikely to be informative about other forms of cheating that may involve students (like the use of cheat sheets and impersonation) or teachers (e.g. giving students extra-time or teaching to the test).

There are no sanctions on either principals, teachers or students suspected of cheating. The individual exams that are flagged as suspicious of cheating are not taken into account in the estimation of the school score that is reported in the official results. The Federal Secretary of Education delivers to the State Ministries a report with the list of the schools in which a high prevalence of cheating is detected.

3.2 School and Locality Characteristics

I use the census of schools carried out by the Secretary of Education (Formato 911) to obtain annual information about school inputs (school and class size, student characteristics and teachers' credentials). Using the census locality code, I retrieve information from the 2010 population census about the characteristics of the localities where the schools are located and from the National Commission for the Evaluation of Social Policy about the localities' poverty rate. I obtain from Google Maps the estimated travel distance by car from the schools' localities to the State capital.[15]

3.3 Census of Teachers

I benefit from extensive data of school personnel compiled as the result of a recent mandate of the Mexican Federal Congress. The data comprises the quarterly payrolls of public elementary schools from the 2nd quarter of 2010 (the last of the academic year 2009-2010) to the 2nd quarter of 2011. The Federal Secretary of Education (SEP) assembled the dataset using information supplied by the State Education Ministries. I track teachers through schools and quarters using their taxpayer number and construct a quarterly panel of school personnel inclusive of name, tax payer and population identification numbers, birth date, assigned school(s) and occupation information. The dataset does not include complete information about hiring, education profile or assigned classrooms.

I do not observe directly in the data who the new teachers hired since 2008 are –since the test-based examination was implemented. However, I can use the 2009 and 2010 censuses to identify the 2010 cohort of new teachers. Then, I match these observations to the list of test-selected teachers, which is available in the dataset. Hence, I focus my analysis on the (24) states that opened vacancies for the Telesecundaria system in the 2010 test-based hiring.[16]

[15] Using the Stata command traveltime.

[16] 2 states (Michoacan and Oaxaca) do not participate at all in the test-based examination and 6 states did not open to

I identify the 2010 cohort of new teachers by comparing the census of Telesecundaria's personnel for the 2nd-quarter of 2010 (the last of the school year 2009-2010) to the census of all personnel registered in any of the four quarterly censuses of the 2010-2011 academic year. I assume that all the 2010-2011 observations that I do not find in the 2nd quarter of 2010 correspond to new personnel in the 2010 school year. I drop observations from the state of Guerrero because there are large missing values for school assignment in the 2nd-quarter of 2010. I also drop 4 states that report relatively few personnel in the second quarter and hence have a high, and likely unreliable, ratio of new/total personnel in the 2010 school year.[17]

The SEP dataset includes a module with the list of the 2009 and 2010 test-selected applicants (573 and 492 teachers, respectively, in the 19 states). I merge this module with the main dataset using the national population number.[18] I am able to merge 76.8% of the 2009 and 89.4% of the 2010 test teachers to specific schools. [19] I obtain the test scores of all the test-selected teachers in the sample by merging the data by full name with the official results available on the website of the Secretary of Education.

I find that 15.5% of the matched individuals hired in the 2010 test examination as new teachers were already in a Telesecundaria's payroll in the 2009 school year. At the extreme, 15 of the 16 test teachers hired in the state of Nuevo Leon fall into this case. This evidence suggests that some incumbent teachers, maybe hired under temporary contracts, where allowed to participate in the examination for new teachers. As I am interested in studying the performance of teachers who are effectively new, I drop the observations from incumbent teachers hired as new teachers in the test examination as well as all the observations from the state of Nuevo Leon and the Federal District.

I can identify, using a module in the dataset, all those who applied for the test-based selection. I do not observe the year and the teaching position of application nor the test result.

The database is inclusive of teachers, administrative staff and principals. I identify as teachers all observations for which I observe, at least in one quarterly database, a synonymous or abbreviation of the word "Teacher" or "Hours Telesecundaria" in the two variables with information about the post description.

I collapse the teachers dataset at the school-year level and merge it into the panel with school results and characteristics. I obtained from the Federal Secretary of Education a list with the schools where the 2008 test teachers were initially assigned (the file does not include the teachers' population or tax identification numbers). I add this information to the panel of schools.

I restrict the dataset to schools which have never received a 2009 or 2008 test teacher, which did not

test-based recruitment any vacancy at the Telesecundaria system in 2010 (Baja California Sur, Colima, Nayarit, Queretaro, Sonora and Zacatecas).

[17] These states are Baja California, Tabasco, Veracruz and Yucatan.

[18] The taxpayer identification number is not available in the test hiring module. Around 5% of the observations in the main personnel module have missing information for the national population number.

[19] I add a smaller number of observations from applicants who were put in waiting list for hiring (16 in 2009 and 68 in 2009). Not surprisingly, the matching rate for this group is considerably lower (50% for 2009 and 15% for 2010)

receive both a new test teacher and a new regular teacher in 2010 and for which there is at least 4 years of Enlace results. I finally trim schools with at least one year-to-year change in their school score larger than 2 standard deviations, which roughly corresponds to the top and bottom one percentiles of score changes. I obtain a dataset with 1,427 schools in 15 states after these restrictions. 6.7% of schools in the sample received, at least, a test teacher in the school year 2010.[20]

4 Allocation of Teachers to Schools

The 2008 reform introduced (within-state) variation in teacher hiring to the education system. Using this variation to meaningfully compare student outcomes in schools that receive test versus discretionary teachers requires, however, that the matching between schools and teachers of the two types is independent of potential outcomes. In a difference-in-differences framework, such a requirement would be violated if state authorities allocated test(discretionary) teachers to schools that would have improved with respect to the comparison schools even in the absence of test teachers –because, for example, the outcomes in these schools were already in a better trajectory before the allocation of test teachers. Figure 1, presented in the introduction, shows evidence against this case, and also indicates that test teachers are allocated to schools that have on average lower levels of student achievement. I proceed now to study the institutional setting that governs the allocation of teachers to schools and generates such patterns.

The allocation of teachers to schools is centralized at the state level and is run by joint committees of state officials and union representatives. Joint committee decisions are mandatory for education officials and school principals. The allocation is done in two steps. First, school vacancies are open for application among current teachers, which leads to sequential inter-school transfers, as the position left by a teacher who fills a vacancy is open for application to teachers from other schools. This process, known as *el corrimiento*, stops when no current teacher is interested in the available school positions. Then, new teachers are assigned to these schools.

The allocation of current teachers (inter-school transfers) is highly regulated and depends on teachers' preferences and committee evaluation of their merits.[21] Typically, as teachers progress in their careers they move from schools based in more isolated and poorer localities to schools in more urban and wealthier localities. The allocation of new teachers (both test and discretionary-hired) does not have to follow the same criteria and joint committees enjoy more discretionary power in this process. In principle, one could expect no difference in the characteristics of schools where test and discretionary teachers are allocated. However, such differences could arise if teachers prefer localities with better amenities, and union representatives are able to influence committee decisions and give a higher weight to the preference

[20] I exclude all observations from the states of Morelos and Campeche because there is no left schools with only test teachers after these restrictions.

[21] Joint committees operate under state-level regulations heavily based on a 1973 agreement between the Federal Secretary of Education (SEP) and the teachers' union (SNTE). Committees must evaluate candidates according to their schooling, tenure, ability and discipline.

of the (new) discretionary teachers. The data is consistent with this hypothesis.

Table 2 shows descriptive statistics for the set of Telesecundaria schools that received either test (treatment) or discretionary (control) teachers hired in the 2010 school year. [22] Test teachers tend to be allocated to schools with lower performing students, headed by a principal with less schooling and based in localities which are poorer, further away from the state capital and have lower penetration of public services.

Specifically, students in schools where test teachers are allocated have, pre-treatment, lower scores in mathematics (by .22 national standard deviations) and Spanish (.19 SD) –in the Enlace standardized exam– than those in schools with new discretionary teachers; and are more likely to be of indigenous origin (an indicator for poverty) (16% versus 10% , though this difference is marginally insignificant). Consistently with these differences, treated schools are based in localities which are on average around 45 minutes further away from the state capital and have higher poverty rates (by around 11 percentage points). Differences in school inputs are less clear. Class size is similar (19-19.7) in the two groups of schools, but schools with test teachers are less likely to have a principal with graduate school training (20% versus 32%).

As schools' location, inputs and outcomes are all correlated, a regression analysis can be more informative about the process generating the allocation of teachers to schools than mere binary comparisons. Hence, I estimate a linear probability model in which the dependent variable is an indicator that turns 1 if the school received a test teacher in the school year 2010 (treatment) and 0 if it received a discretionary teacher (control) in the same year. I regress this indicator on a vector of (past) school outcomes, inputs and locality characteristics, plus state fixed-effects. Table 3 reports the results.

Holding constant school inputs and locality characteristics, no single measure of student achievement in the last two years predicts assignment into treatment. Moreover, the p-value associated with the test of the joint significance of all the (past) student-performance variables included in the model is high (.265); and I cannot reject the null hypothesis that they are jointly insignificantly different from zero at conventional levels of statistical significance. I obtain similar results if I use rates of change instead of levels for measuring past school performance. So, the data does not support an assignment model in which joint committees allocate test teachers based on past school performance.

In contrast, the strong statistical relationship between treatment status and locality characteristics shown in the bivariate analysis remains after conditioning on student performance and school inputs. Test teachers are more likely to be allocated to schools in localities which are poorer, have a lower penetration of electricity and are further away from the state capital. Again, the relationship between treatment status and school inputs is less startling. The allocation of a test teacher has a positive and significant partial correlation with class size (though the magnitude of the coefficient is small .00263) and negative

[22] I focus only in schools receiving new teachers and exclude here and after schools that receive both test and discretionary teachers and schools that received test teachers hired in 2008 and 2009.

and insignificant with principal's education.

Overall, the regression analysis indicates that treatment status is strongly correlated with locality characteristics and, to a lesser degree, with school inputs. Also, there is no observed relationship between past school performance and the probability of assignment into treatment, once locality characteristics are taken into account. These results are consistent with the hypothesis that committees allocate test and discretionary teachers based on teachers' preferences for locality characteristics and not on past school performance and in which committees give a higher weight to the preferences of the discretionary teachers. This is encouraging evidence for a difference-in-difference analysis, in which it is possible to control for both the effect of time-invariant locality characteristics and time-variant (observable) school inputs.

5 The Effect of Test-Hired Teachers

5.1 Identification and Estimation Methods

I am interested in estimating the average effect on student outcomes of assigning to a school a new teacher selected in a test-based examination (treatment) versus assigning a new teacher selected by the teachers' union in a discretionary process (control). My identification strategy takes advantage of an allocation process of teachers to schools that is not based on past school trends. With this purpose in mind, I estimate the following difference-in-differences model with school fixed-effects:

$$y_{st} = \beta_0 + \beta_1 share_{st} + \Gamma X_{st} + \tau_t + \alpha_s + v_{st} \tag{1}$$

Where y_{st} is an outcome of school s at time t, $share_{st}$ is the share of 2010 test teachers among total teachers in school s at time t, β_1 is the parameter of interest, X_{st} is a vector of time-variant school inputs, Γ is the associated vector of parameters, τ_t is a vector of year effects fully interacted with state dummies, α_s is a school time-invariant component and v_{st} is a disturbance term.

I estimate the model using panel data of Telesecundaria schools that receive new teachers, either test or discretionary-hired, in the 2010 school year. I focus on Telesecundarias because their small size should increase the likelihood that I can statistically observe a teacher effect at the school level. I approximate school outcomes with 9th-grade outcomes in my main estimations. I do so because I observe 9th grade test scores for a longer period than 8th and 7th grade test scores, and hence I can study and control better for pre-treatment trends in outcomes. [23]

The model in all regressions controls for the number of 2010 new teachers in the school, class size, school size, the share of indigenous students, principal's attendance of graduate school and a vector of

[23] I have 9th-grade test scores from 2005 to 2010 (five years before and one year after treatment), while I only have 7th and 8th-grade test scores from 2008 to 2010.

interactions between year and state dummies to capture state-specific time trends. Standard errors are clustered at the school level.

Under the parallel trends assumption, the difference-in-differences parameter β_1 captures the total (average) effect of increasing the share of test teachers in a school from 0 to 1. Or in other words, β_1 measures the difference in the (total) value-added (teacher quality) to student outcomes between test and discretionary teachers.

I model the relationship between teachers and students at the school level for identification, though there are reasons for which one could be interested in estimating the average treatment effect at the classroom level. For example, teachers may not teach all students in a school. In Telesecundarias, teachers actually teach only one classroom per school. So, the classroom might be a more natural unit to conceptualize their influence. I cannot directly link teachers to students in the data though. But even if I could, the matching between teachers and students and the potential within-school externalities of teacher quality make the identification of a causal effect at the classroom level more restrictive.

I describe and empirically investigate the allocation of teachers to schools and find strong support for the parallel trend assumption necessary for the identification of a causal effect at the school level in a difference-in-differences model. I cannot do the same for the process generating the within-school allocation of teachers to students. Even when more detailed data is available, Rothstein (2010) gives a critical assessment of the typical assumptions made about the assignment of students to teachers in which observational studies rely on to identify teacher causal effects.[24]

The focus at the school level allows me also to neglect the within-school externalities associated with teacher quality. More effective teachers could, for example, free up other school resources, like a principal's time, for the benefit of students in other classrooms. Also, better teachers might have a direct effect on students in other classrooms, for example, through personal interactions.

My specification of treatment intensity provides a scaled-up treatment effect. As said before, β_1 captures the total (average) effect of increasing the share of test teachers in a school from 0 to 1. So, given a one-to-one relation between teachers and classrooms in a school, β_1 can also be interpreted as the average effect of a test teacher on student achievement in a classroom.

5.2 Past Trends in Outcomes

The causal interpretation of β_1 requires that the control schools give an accurate counter factual of the outcomes that the treated schools would have had in the absence of treatment. Although it is impossible to test directly this assumption, I take advantage of observing school outcomes for five years before treatment and test whether the trends in both sets of schools were the same during the pre-treatment period.

[24] Although the Rothstein's critic is focused in the estimation of individual teacher effects which require stronger assumptions than the estimation of an aggregated effect.

Figure 1 shows the raw data, presenting the evolution of mean school scores by (eventual) treatment status. The visual evidence is encouraging. Schools that receive (new) test teachers in 2010 have on average lower school scores, pre-treatment, than schools that receive (new) discretionary teachers. But crucially for my identification strategy, the outcomes of both sets of schools seem to follow similar paths during the pre-treatment period, and then converge after treatment (year 2010).

More formally, I estimate a modified version of equation 1 in which I regress the outcomes under study, in separate regressions, on a vector of year dummies interacted with eventual treatment status, plus the set of time-variant school inputs, state-specific time trends and school fixed-effects. I use only observations from the five years in the pre-treatment period. Table 4 reports the results.

In the same line that results in Table 3 and Figure 1, I do not find a consistent statistical relationship between treatment status and the the pre-treatment path of the four outcomes that I study: enrollment at the end of the academic year,[25] the share of students suspected of exam cheating and the scores for mathematics and Spanish in the Enlace exam. Only one of the 16 estimated coefficients for the interactions between eventual treatment status and year dummies is statistically significant.

The interaction between eventual treatment status and the year 2008 in the regression for final enrollment (column 1) is highly statistically significant and has a large magnitude (10 percentage points). However, the other 3 coefficients in the same regression have small magnitudes (less or equal to 1.4 percentage points) and are not statistically significant. In the same line, no coefficient of interest in the regressions for exam cheating, mathematics and Spanish scores (columns 2 to 4) is statistically significantly different from zero at conventional levels. The magnitude of the coefficients for exam cheating goes from very small to relatively large (magnitudes are .02, .08, 3.1 and 5.4 percentage points from smallest to largest). For mathematics and Spanish, 3 of the 8 coefficients of interest have relatively large magnitudes (around .10 to .15 national standard deviations). But, again, in no case is a coefficient statistically significant at the 10 percent-level. Furthermore, all coefficients have a negative sign. Hence, if anything test teachers would be allocated to schools with a negative trend in test scores, which if true would produce a downward bias in my estimates.

Overall, I cannot reject the null hypothesis that the pre-intervention year dummies are the same for both control and treatment schools at conventional levels of statistical significance in three of the four regressions. I interpret these results as strong evidence in favor of the parallel trend assumption necessary for the identification of a causal effect in a difference-in-differences model.

5.3 Main Results

I present the main results of my difference-in-differences estimation in Table 5. I evaluate first if the allocation of (new) test teachers changes the enrollment rate at the end of the academic year (see column

[25] Measured as the number of students that take the Enlace exam over the number of students registered at the beginning of the academic year.

1), and I do not observe any effect. The coefficient of interest has a negative sign, but the magnitude is very small (1.3 percentage points) and the point estimate is not different from zero at conventional levels of statistical significance.

In contrast, I find that test teachers considerably decrease exam cheating (column 2). Going from not having test teachers in a school to having only test teachers is associated with a reduction of 5.8 percentage points in the share of students suspected of exam copying, a result that is statistically significant at the one-percent level. This is a large effect: it amounts to the prevalence of exam cheating in the year before treatment.[26]

In a seminal study, Jacob and Levitt (2003) detect an increase in teacher cheating in student exams in response to incentives that reward high test scores in Chicago Public Schools. More recently, Angrist, Battistin and Vuri (2014) find that teachers cheat to reduce the effort of grading open-ended questions included in standardized exams in the south of Italy. Closer to the context of this study, Behrman, Parker, Todd and Wolpin (2015) find an increase in exam cheating after the introduction of monetary rewards for students, teachers and principals in a sample of senior high schools in Mexico (using a different test, not Enlace).

As explained before, students have no explicit incentives to perform better in the Enlace exam (and benefit from exam cheating), as test scores have no bearing in graduation or GPA. In contrast, teachers can receive monetary bonuses for high Enlace scores. Also, the accountability setting around Enlace can put pressure on teachers and principals to perform better. The estimated results suggest that the (new) discretionary teachers are more willing or able to react to these incentives by engaging in or facilitating exam copying.

Turning to the main outcome of interest, the allocation of test teachers has a positive impact on student achievement in mathematics and Spanish (columns 3 and 4), which is both statistically and economically meaningful. The treatment coefficients in the achievement regressions have a large magnitude (.52 national standard deviations for mathematics and .31 national standard deviations for Spanish), and are statistically significant at the one-percent level in the two cases.

The estimated gains are larger than those in related studies of teacher hiring. Duflo, Dupas and Kremer (2015) find that hiring contract teachers to reduce class size (from 82 to 44 students) increases student achievement by around .23-.25(.17-.19) standard deviations in mathematics(Literacy) in a sample of Kenyan primary schools –the authors find evidence that suggests that their results are not driven by the reduction in class size. In contrast, Muralidharan and Sundararaman (2013) estimate that contract teachers produce the same value-added that unionized teachers in primary schools with –experimentally induced– smaller classes in India. In both studies though, the authors compare new to experienced

[26] As explained in Section 3.1.1, the Federal Secretary of Education measures exam cheating using a detection algorithm designed to give lower-bound estimates for direct copying, e.g. one student copying from another or a larger group of students (and potentially teachers) exchanging responses during the application of the exam.

teachers, which is important to consider if, as the literature indicates, teachers produce less value-added during their first years of experience. In the United States, Kane, Rockoff and Staiger (2008) find no or small (.02 SD) differences in value-added when comparing new teachers selected using four different certification paths in New York City public schools. More tellingly, the authors estimate that the value-added by mathematics teachers in the top quartile of the value-added distribution is .33 SD larger than the one produced by those in the bottom quartile.

In perspective, the gains produced by test teachers are similar to those in Behrman, Parker, Todd and Wolpin (2015), who evaluate an ambitious incentive scheme to boost student achievement in mathematics in Mexico. In randomly selected schools, students, teachers and principals receive monetary rewards linked to test scores at a cost per student of 15% of the total per student expenditure in Federal senior high schools. Accounting for a related increase in exam cheating, the authors estimate that the monetary incentives produce an increase of .42-.63 standard deviations in test scores in the second year of program implementation.

Summing up, the gap in the value-added by test and discretionary teachers is larger than the one estimated between the top and bottom 25-percent teachers in New York elementary schools and is similar to the gains produced by an ambitious, and costly, incentives program set up in a closer context. So, the test-based selection leads to the hiring of teachers who are much more effective than those hired through the discretionary process.

5.4 Robustness Checks

5.4.1 Spurious Outcomes

Telesecundarias cater for small and isolated villages where local youth face high transportation costs to attend a regular junior high school located in a larger locality. Given the limited school choice, I do not expect to observe that the allocation of test teachers changes the composition of students in a school at the beginning of the academic year. A large correlation between the allocation of test teachers and the composition of students in the school would indicate the presence of underlying differential trends between treatment and control schools that the test on past outcomes fails to detect, but that could bias the reported results. With this purpose in mind, I run my main model using as outcomes the share of indigenous students and repeaters in the school. Table 6 reports the results.

I do not observe any correlation between the allocation of test teachers and both the share of indigenous students and repeaters in the school (see columns 1 and 2, respectively). The point estimates of interest have a small magnitude (3.7% and .02% respectively) and are not statistically significant at conventional levels.

5.4.2 Contemporary Shocks in Other Inputs

I also estimate my main model using as outcomes two measures of school inputs: an indicator for whether the principal has a graduate education and class size. The idea is to investigate the presence of potential contemporaneous shocks that could introduce bias in the main results. The concern would be that for some policy (about which I do not have any information) the schools where test teachers are allocated experience at the same time an improvement in (other) school inputs. Results are, again, in Table 5.

Both the coefficients of interest for the principal's education (column 3) and class size (column 4) have a negative sign and small magnitude (3.9 percentage points and .04 students). The point estimate for the principal's education is marginally significant, while the one for class size is not statistically significant at the ten-percent level. So, I do not find evidence of contemporaneous positive shocks in other inputs in the schools that receive test teachers.

5.4.3 School Time Trends

For further robustness checks, I modify the main equation to allow each school to have a specific linear time trend. Hence, I run first separate regressions of every outcome on a full set of school dummies interacted with a linear time trend (using the five pre-treatment years) and use the residuals to predict (for the whole time period) a de-trended outcome. Then, I take the de-trended outcomes to my main specification. Panel A in Table 7 reports the results.

The same story as before emerges from this estimation. There is no effect of test teachers on enrollment at the end of year (column 1), but there is a large effect on the probability of cheating (columns 2) and student achievement (columns 3 and 4). Point estimates have both a magnitude and statistical significance similar to those shown in Table 5.

5.4.4 Binary Treatment

Up to now, I have modeled treatment as the share of test teachers in the school with the purpose of providing a scaled up estimate of a teacher effect in a regression run at the school level. One could be worried though that a few schools with a high share of test teachers drive the results reported so far. Hence, I produce estimates using a binary specification for treatment: whether or not (at least) a test teacher is allocated to the school. The counter factual is again the allocation of a newly-hired discretionary teacher(s). I restrict the estimation to small schools, those with 3 classrooms or less, as it is highly unlikely that I can find a statistically significant effect in large schools. Panel B in Table 7 reports the results which, again, are reassuring.

No changes in the conclusions arise from this set of estimates: the allocation of test teachers does not have an effect over enrollment at the end of the academic year, but it reduces suspected cheating and increases student achievement, measured by the mathematics and Spanish scores in the Enlace exam.

5.4.5 Heterogeneous Response

The causal interpretation of the difference-in-differences parameter β_1 relies on the assumption that the average outcomes of the control group follow the same trend that the average outcomes of the treatment group would have followed in the absence of treatment. In support of the so-called parallel trend assumption, I showed in Table 4 that treatment and control groups followed similar paths in the four outcomes under study during the five years previous to treatment. However, another source of concern arises from the potential interaction between the outcome variables and exogenous characteristics with unbalanced distributions between schools with new test and discretionary teachers (Meyer, 1995; Heckman, Ichimura and Todd, 1997). For example, if the impact of test teachers is larger in isolated localities.

Abadie (2005) proposes a semi-parametric difference-in-differences estimator to deal with non-parallel outcome dynamics for the treated and control groups due to differences in observed characteristics. The estimation uses a two-step strategy in which first a propensity score is estimated and then a matching estimator re-weights the control observations on the propensity score and imposes in a common support a balanced sample in pre-treatment characteristics between the treated and the control groups.

Hence, I estimate first the propensity score from a logit model of the probability that a school receives a test teacher in the school year 2010 as a function of similar pre-treatment characteristics to those listed in Table 3 plus two lags of final enrollment and exam cheating. The model includes third-order polynomial functions for all the (continuous) variables. I impose a common support by dropping both treatment observations of which the propensity score is higher than the maximum propensity score of the control observations, and control observations of which the propensity score is lower than the minimum propensity score of the treated observations. The Abadie's estimator matches differences in pre-treatment and post-treatment outcomes for the treated to weighted averages of differences in pre-treatment and post-treatment outcomes for the untreated. Hence, I use only observations for one year before and one year after treatment. I use a binary definition of treatment and restrict the estimation to small schools (those with 3 classrooms or less). Results are reported in Panel C on Table 7.

The main results are consistent with those in Panel B on the same table. The allocation of a test teacher does not have an discernible effect on the share of students in the school that take the Enlace exam at the end of the academic year (column 1). The difference is that I do not observe a statistically significant effect on the share of students flagged for copying in the final exam (column 2). However, in the same line that in Panel B, I observe a positive effect on mathematics scores of .28 national standard deviations, which is statistically significant at the one-percent level and a positive effect on the Spanish score of 16.7 national standard deviations, statistically significant at the five-percent level.

5.4.6 Classroom cohorts

I have run all regressions in a panel of 9th-grade scores because it allows me to study and control for pre-treatment trends in outcomes using a larger time dimension. The application of the Enlace exam started in 2005 sitting only ninth graders, but it also covers students in 7th and 8th grades since 2008. Hence, alternatively, I can use the outcomes of students in 8th and 7th grade in 2009 and 2008, respectively, the two years prior to treatment, to investigate the robustness of my results to the choice of pre-treatment outcomes.

Figure 2 shows the raw data and Panel A on Table 8 the falsification test to investigate differences in pre-treatment trends correlated to eventual treatment status. I do not find any correlation between the allocation of test teachers in 2010 and changes in student outcomes between 2008 and 2009. The coefficients for the partial correlation between eventual treatment status and changes in enrollment at the end of the academic year (column 1) and suspected cheating at the Enlace exam (column 2) have a very small magnitude (.04 and 1.5 percentage points, respectively) and are not statistically significant at conventional levels. The coefficient of interest is relatively large in the regression for mathematics score (.17 SD, see column 3), but it is not statistically significant at conventional levels and the sign is negative. The magnitude of the corresponding coefficient in the regression for Spanish score is smaller (.05 SD, see column 4) and is not statistically significant. This test gives additional support for the identification of a causal effect based in the parallel trend assumption.

The results, available in Panel B on Table 8, are in line with those presented before. First, there is no observed effect on enrollment at the end of the academic year. The coefficient of interest has a small magnitude and no statistical significance, see column 1. Then, there is a significant effect on suspected cheating (negative, see column 2) and student achievement (positive, see columns 3 and 4), measured by both mathematics and Spanish scores. The coefficients of interest confirm that test teachers have a large and economically significant effect on student outcomes, as those reported in the main specification (Table 5).

5.5 Distributional Effects

The results presented so far show that allocating to a school a test teacher over a discretionary one raises considerably the mean of student achievement. It is possible though that this average effect hides a pattern in which the two group of teachers tend to focus on different types of students. For example, discretionary teachers could lean more towards the welfare of relatively low-achieving students (within a classroom) and focus the teaching on these, while test teachers focus the teaching on high-achieving students. If learning gains depend strongly on the level of past achievement, such specialization could produce the average effect observed, with important implications for the interpretation of this result. I proceed now to investigate such a possibility.

I run my main specification using as outcomes –instead of the mean school scores– the 25th and 75th percentiles of the mathematics and Spanish score distribution in each school. Table 9 reports the results. The effect of test teachers in both mathematics and Spanish achievement is similar at the 25th and 75th percentile of the score distribution, and mimics the results presented in the previous subsection. This evidence does not support the hypothesis that the observed average effect is due to a specialization of each type of teacher on students of different characteristics. Both high and low-achieving students in a school benefit from the allocation of a test-hired teacher. [27]

6 Why Discretionary Teachers Do Worse

I have presented several tests and specifications supporting the identification of the causal effect of allocating to a school a new teacher hired through the test examination versus a new teacher hired through the discretionary process. Results show that test teachers do considerably better. The estimates in my main specification indicate that going from having zero test teachers in a school to having only test teachers increases student achievement by .52 standard deviations in mathematics and by .31 standard deviations in Spanish. This is an important finding for policy. Using the test leads to the hiring of considerably more effective teachers than relying on the discretionary process.

A related policy question is to understand the mechanics that lead to such a quality gap. To do so, it is useful to decompose the total effect of using a recruiting method over the other in: 1) the difference in the efficacy of each method at selecting teacher quality given a pool of applicants; and 2) the difference in the pool of applicants that each method attracts. I turn now to study, first, the efficacy of the teacher test at identifying teacher quality.

6.1 Teacher Testing and Teacher Quality

6.1.1 Does the Teacher Test Predict Teacher Quality?

There is a large literature focused on predicting teacher quality using teacher test scores. Hanushek and Rivkin (2006) review the literature and observe that teacher scores have more frequently been significantly correlated with student outcomes than other characteristics like education, experience (beyond the initial years) and salaries. However, they also characterize the evidence as far from overwhelming. The authors identify a set of nine high-quality studies from which six find insignificant results, two find significant positive results and one finds significant negative results.[28] A similar panorama emerges from a review of U.S. studies conducted by Wayne and Youngs (2003). They identify a set of seven studies from which five find positives results and two find negative results.

[27] I also test and find that the schools that receive test and discretionary teachers experienced parallel trends in these outcomes during the pre-treatment period. Results are available by request.

[28] The authors define high-quality studies as those which use value-added estimation where the sample is drawn from individuals from a single U.S. state.

It is frequently thought that tests that measure knowledge of the teaching subject do better at predicting teacher quality than other tests. The findings of Glewwe, Hanushek, Humpage and Ravina (2011) in a literature review of studies in developing countries are consistent with this idea. The authors identify a set of five studies about knowledge of the teaching subject from which four find significant positive results and one finds positive insignificant results.

Hanushek and Rivkin (2006) point out that even when significant, teacher test scores capture only a small portion of the overall variation in teacher quality. The results of three studies which estimate value-added models using rich panel data from students and teachers illustrate this claim. Goldhaber (2007) (using North Carolina data) finds that moving from a teacher score in the bottom quintile of the test distribution to a teacher score in the top quintile of the test distribution leads to statistically significant increases in student achievement of around .009-.013 standard deviations in reading and .037-.03 in mathematics. Clotfelter, Ladd and Vigdor (2006)(also using North Carolina data) find that a 1-SD increase in teacher test performance increases fifth-grade achievement by .01-.02 SD. Finally, Rockoff, Jacob, Kane and Staiger (2011) (using New York City public schools data) find that a 1-SD increase in mathematics knowledge for teaching increases by .03 SD in achievement in mathematics (a result that is marginally significant). They also find a similar quantitative result for a teacher pre-screening test (the Haberman Prescreener) widely used in the United States (though in this case the estimated coefficient is marginally insignificant).[29]

Summing up, even when statistically significant, which is more often the case when they measure subject knowledge, teacher test scores are modest predictors of teacher quality, as the magnitude of the reported coefficients indicate.

6.1.2 Teacher Test Scores: Identification

I am interested in the predictive power over student achievement of the specific test used to select teachers in the competitive examination. As said, the test is designed to measure cognitive skills, subject knowledge, mastery of teaching methods and ethics in teaching.[30]

I estimate the following value-added model in which I control for past levels of achievement in the school using a fixed-effects estimation:

$$y_{st} = \beta_0 + \beta_1 tscore_{st} + \Gamma X_{st} + \tau_t + \alpha_s + v_{st} \tag{2}$$

Where $tscore_{st}$ is the (mean) test score of the 2010 test teachers allocated to school s at time t, β_1

[29] Metzler and Woessmann (2012) use data from rural Peru and find a larger effect (.1 SD) of knowledge of teaching subject on student achievement than the cited studies. Their identification strategy is based on the within-teacher within-student variation of test scores in two subjects and hence their estimates suffer from upward bias if the variation between scores is related to something else that knowledge. For example, if teachers have higher motivation in subjects in which they are more knowledgeable.

[30] These are the exam contents for the 2010 exam. The contents measured in the exam changed over the years.

is the parameter of interest and all the remaining terms are the same as in equation 1. I include in the vector X_{st} the share of 2010 test teachers in school s at time t, plus the same time-varying school inputs as in the achievement model. I restrict the estimation to schools that received test-hired teachers as I only observe teacher scores for this type of new employees. I run specifications using past 9th-grade scores and 8th-7th grade scores as measures of past achievement. Standard errors in all regressions are clustered at the school level.

If test scores capture teacher quality, then teachers with higher test scores should significantly produce higher student achievement. The identification assumptions are that the cumulative effect of the entire history of students' past inputs is captured by the past achievement levels and that the allocation of test teachers to schools is orthogonal to potential learning gains.

As said, I can only estimate equation 2 in schools that receive test-hired teachers and that are in my main sample. The reason is that only in this set of schools can I observe both student outcomes and teacher test scores. This restriction reduces the sample size, affecting the statistical power necessary to identify an effect that is different from zero. The small sample size is particularly important because the literature review indicates that the size of the coefficient of interest is potentially small.

The restriction is the result of a process truncating the data based on the main independent variable: teacher test scores. I do not observe the school assignment of applicants with scores below the state's hiring threshold. The reduction in the variation of teacher test scores affects mechanically the precision in the estimation of the coefficient of interest, and can also introduce bias if the relationship between teacher test scores and teacher quality is not linear.

Figure 3 presents the distribution of teacher test scores. Panel A shows the distribution for all applicants in the test examination (for Telesecundaria positions in 2010 in the 15 states in the sample). The red line indicates the (approximate) threshold below which applicants are not considered fit for a teaching position. The Federal Secretary of Education defined this threshold as 30% of total items in the test. Only one percent of applicants scored below this certification-like threshold.

Panel B presents the distribution of teacher test scores in the schools in the final sample. Individual test scores are aggregated at the school level because some schools receive more that one test teacher. A mere glance indicates clear differences between the two distributions. Not surprisingly, the mass of the distribution in Panel B is skewed to the right. However, there are observations along almost all the support of the test scores distribution, or at least in the range above the certification cutoff. This is because there is variation in the hiring cutoff at the state level. Some states hired a large number of applicants pushing down the hiring cutoff.

6.1.3 Teacher Test Scores: Results

Table 10 presents the results for the estimation of equation 2. Columns 1 and 2 report results for mathematics and Spanish achievement for the specification with past 9th-grade scores as a measure of pre-treatment achievement. The coefficient of interest is positive in both cases, but not statistically significant at conventional levels. The magnitudes imply that an increase of 1 SD in the teacher score is associated to a decrease of .005 standard deviations in student achievement in mathematics and an increase in .04 standard deviations in Spanish. Again, both coefficients of interest are imprecisely estimated and are not statistically different from zero at conventional levels of statistical significance.

Columns 3 and 4 report results for the specification with achievement in the past two grades (8th and 7th) as a measure of previous student achievement. Results are not robust and now the sign for the effect on mathematics is positive, while the one for Spanish is negative. Again, the coefficients of interest have a small magnitude and are not statistically significant at conventional levels.

The results in Table 10 indicate that the exam used to select new teachers has no predictive power over teacher quality. Teachers with different test scores have the same effect on student achievement. Ruling out non-linearities in the relationship between teacher and student test scores, the expected quality of those selected for their higher test scores corresponds to the expected quality of those who applied to this process.

As said, the estimation of equation 2 has limitations. It is reassuring though that results are in line with the findings in the literature.

6.2 Does Selection to Application to Test-Taking Matter?

Results so far show that the teachers' union selects for hire applicants of a considerably lower quality than those selected using a standardized test with no predictive power over teacher quality. Given this puzzling result, it is natural to ask whether the quality of the test and discretionary applicants differs; and, if this is the case, if differences in applicant quality can explain the large quality gap between those hired through each method.

The test and discretionary recruitment have different institutional characteristics, that can likely lead to differences in the distribution of applicant quality by selection method. However, it is not obvious to expect either the test or the discretionary applicants to be on average of higher or lower quality than their counterparts, as there are factors which might shape selection to application in opposite directions.

The test-based recruitment restricts applications to Telesecundaria positions to graduates from teacher-training schools (which are tertiary-level institutions). In contrast, the discretionary process is open to both graduates from teacher-training schools and from other undergraduate institutions. So, in principle, the teachers' union can select from a larger and more diverse pool of applicants.[31]

[31] Depending on the state, graduates from a set of specific fields of study can apply to other teaching positions.

Angrist and Guryan (2008) highlight that testing imposes a cost to the applicant that shifts the labor supply to the left. If the cost is common to all applicants, testing could have a stronger deterrent effect on higher-quality applicants (those with a larger outside option and hence on the margin on the decision to apply or not). However, testing could further decrease the applications of lower-quality applicants if the cost of studying for the test is negatively correlated with teacher quality.

Applicants to the discretionary process could face a different cost. Testing was implemented amid large complaints that applicants to the discretionary process would have to pay bribes to be hired. Such side payments would shift supply to the left and particularly deter applications from higher-quality individuals.

Employment conditions do not depend in principle on the hiring method. Salaries follow a pay scale defined at the state level and new teachers get tenure after three months in employment. However, the analysis in Section 4 reveals that discretionary-hired teachers tend to be allocated to schools based in localities with better characteristics (closer to the state capital, less poor and with higher coverage of public services). So, if applicants value the characteristics of the localities where schools are located, application through the discretionary method has on expectation better non-wage employment conditions. Higher expected benefits should shift supply to the right, with a stronger effect among high-quality applicants.

Summing up, whether there are –or not– differences in applicant quality across methods is an empirical question. Though I do not have information to test directly for this, I can say something about the efficacy of each method given a similar pool of applicants. Specifically, I can compare the quality of test applicants selected by the test and by the discretionary process –as I am able to identify the test applicants who were hired through the discretionary process.

Comparing the efficacy of the two recruitment methods holding constant the pool of applicants is particularly informative because the test bears no relation to teacher quality. So, I can investigate whether the discretionary process is better, equal or worse than a selection rule that produces similar results in terms of teacher quality than random selection. This is a useful exercise to understand if the large quality gap between test and discretionary teachers is fully explained by the selection of lower-quality applicants to the discretionary process, or if even given a similar pool of applicants the discretionary method leads to the selection of lower-quality teachers.

One important caveat is that I do not have detailed information about the test application of those selected through the discretionary process. Specifically, I do not observe the teaching position, year of application nor the test score obtained. So, I must assume that the quality of test applicants is constant along teaching position and year. I will underestimate the efficacy of the discretionary process if the quality of the test-applicants to Telesecundaria positions in 2010 is higher than the quality of test-applicants in other years and teaching positions.

6.2.1 Results

As before, I study first the pre-treatment trends in outcomes in the set of schools of interest. Figure 4 plots the evolution of test scores in schools with new teachers in 2010 by eventual treatment status. The sample includes the schools which receive either test teachers or discretionary teachers who were also test applicants. Visual inspection suggests that outcomes in the two sets of schools follow similar paths during the pre-treatment period, which I formally test and confirm in Panel A on Table 11.

Focusing on mathematics and Spanish scores, I do not find any statistically significant difference in the pre-treatment paths of outcomes when I compare the schools to which test teachers were allocated with those which received discretionary teachers with test participation (columns 1 and 2). This finding is consistent with an allocation process of teachers to schools that is not a function of past school performance.

Panel B on Table 11 reports results. As before, there is a large difference in the effect of test teachers on student achievement over the allocation of discretionary teachers. The coefficients for mathematics and Spanish achievement are very similar in magnitude and statistical significance to those presented in the main results (columns 1 and 2). Students in schools which receive test teachers do better by .55 national SD in mathematics and .32 national SD in Spanish than students in schools which received discretionary teachers.

The analysis in the previous subsection shows no relation between test scores and teacher quality. Nonetheless, one could be worried that if test scores do predict teacher quality the results presented in columns 1 and 2 could be explained in part by the higher test scores of the test-hired teachers. The estimates coming from high-quality studies on the relation between test scores and student achievement indicate that differences in test scores are unlikely to explain the magnitude of the estimated results. Still, I do not have the test scores to include in the model and completely rule out this possibility. However, I observe teacher characteristics that are correlated with test scores, notably the applicant's college GPA. Columns 3 and 4 present results for regressions which control for the college GPA and age of the (2010) new teachers in the schools. Results are robust to these controls.

So, given a similar pool of applicants, the discretionary process leads also to the hiring of considerably lower-quality teachers. The similarity between these results and those from the main estimations indicates that the observed quality gap between test and discretionary teachers is not explained by the self-selection of higher-quality applicants to the test-based examination.

6.3 Discussion

The presented findings show that the discretionary-hired teachers do considerably worse than the test-hired teachers; even though the standardized exam used in the test hiring bears no relation to teacher quality. Furthermore, when comparing the efficacy of the two methods on a similar pool of applicants,

those who self-select to test application, the discretionary method leads also to the hiring of lower quality teachers. The combination of these results indicates that the discretionary teachers are selected from the bottom of the distribution of applicant quality. Then, joint committees of state officials and union representatives allocate the discretionary teachers to schools located in more "desirable" localities –but with similar pre-treatment trends in outcomes. Although striking, this pattern is consistent with the criticisms of the discretionary process for allowing retiring teachers to hand on their position to relatives and the (potential) selling of teaching positions.

As discussed in Section 2.1, a hiring agent that maximizes private rents –rather than teacher quality– has the incentive to hire low-quality applicants. If the agent decides to put a teaching position up for sale (demands a bribe from applicants), the side payment is maximized by selecting the worst applicant, the one who has the lowest outside option in the labor market and hence the incentive to pay the highest bribe. Similarly, for a retiring teacher who maximizes family income the optimal solution is to select the relative with the worst outside option.

6.3.1 Alternative interpretations: evaluation bias and on-the-job incentives

There are two alternative mechanisms that are, in principle, compatible with the empirical results. First, the discretionary process could lead to the selection of lower-quality applicants due to evaluation bias. For example, teachers could hire their relatives because they believe that they are as qualified, or more, than the alternative applicants. Such a possibility is difficult to test directly, though some elements speak against this hypothesis. For starters, it seems hard to believe that the teachers' relatives are as a group at the bottom of the teacher quality distribution (while it is more likely that a targeted selection within this group fits such a description). Also, in its public opposition to test hiring, the teachers' union framed the hiring of relatives as a labor right rather that as an efficient mechanism to select good teachers (Elizondo, 2011). Finally, there are other elements in the participation of the union in the hiring of new teachers that are consistent with rent-seeking and not with evaluation bias. Notably, the favoritism that the joint committees of state officials and union representatives give to the discretionary teachers in the allocation to schools in more "desirable" localities (closer to the state capital, less poor and with higher coverage of public services) (see Section 4).

Second, one might wonder if differences in after-hiring incentives between test and discretionary teachers could produce the observed gap in teacher quality (value-added). As already discussed, the transfer to other schools, an important mechanism for progress in the teaching career, depends on the evaluation of the joint committees of officials and union representatives. Anecdotal information, and the results presented in Section 4, suggests that closeness to the union can help to achieve better school allocations. If the discretionary-hired teachers are more connected to the union, they might need less on-the-job effort to secure promotion. They could also be more involved in union-related activities at the

cost of tasks directed to increase student achievement if they are selected because they have a comparative advantage in such a role. But, on the contrary, test teachers could have the incentive to participate more in union activities to make up for a potential absence of union connections. It is important to note that affiliation and payment of fees to the union is mandatory no matter how teachers are recruited. Hence, it is not obvious if the incentives faced by discretionary and test teachers vary in a way that would lead to differences in the effort exerted in the classroom. I do not observe in-class effort to explore this specific channel, though it is consistent with the general interpretation of a rent-seeking union which deviates resources from the maximization of student achievement.

To explore whether the observed results can be explained by differences in on-the-job incentives, I investigate whether the allocation of teachers to schools with better institutional settings –that represent a specific type of incentive– make a difference in teacher effectiveness. Specifically, I look at the interaction between the allocation of a test over a discretionary teacher and the participation of the school in the program School of Quality (PEC for its Spanish acronym) (36% of schools in the sample participate in PEC). PEC is a federal program directed to improve school governance that distributes small grants (up to 50,000 pesos or 3,200 USD) to schools that commit to a plan to improve school quality. Participation in the program is optional and depends heavily on the school principal. Hence, one can interpret participation in PEC as an indication of having a motivated principal in the school. The hypothesis is that a better school principal could minimize differences in teacher effort and, hence, in value-added to student achievement. However, I do not find evidence that supports this case. I estimate the main model with the addition of an interaction between the share of test teachers in the school and PEC participation. The coefficients for this interaction term are small and not statistically different from zero at conventional levels (results are available on request from the author).

7 Conclusions

In this paper, I take advantage of a unique setting that allows me to compare the quality (value-added to student achievement) of the teachers hired in a discretionary process led by the teachers' union in Mexico to that of teachers hired on the basis of a screening rule. The presented findings show that the discretionary-hired teachers do considerably worse than the test-hired teachers, even though the standardized exam used in the test hiring bears no relation to teacher quality. Furthermore, when I compare the efficacy of the two methods on a similar pool of applicants, those who self-select for the test application, the discretionary method is also found to lead to the hiring of lower quality teachers. The combination of these results indicates that the discretionary teachers are selected from the bottom of the distribution of applicant quality. I then show that joint committees of state officials and union representatives allocate the discretionary teachers to schools located in more "desirable" localities –but

with similar pre-treatment trends in outcomes. This pattern is consistent with the criticisms of the discretionary process for allowing retiring teachers to hand on their position to relatives and of the selling of teaching positions.

The results show how costly the use of discretion in hiring can be. In a principal-agent framework, if the principal cannot punish the agent for hiring low-quality workers (for technological or political economy reasons), then the principal might be better off relying on a hard-to-manipulate rule for hiring, even if such a rule is an imperfect predictor of worker quality or has not predictive power at all. The reform to teacher hiring evaluated in this study promises to deliver large educational and economic benefits in Mexico through the substitution of low-quality by average-quality teachers. Rules predominate over discretion in this context. Extrapolations to other contexts would have to take into account factors like the principal's capacity both to detect ex-post worker quality and punish(reward) the agent for hiring low-quality applicants; and the degree of competition in the product market (as perfect competition makes any gap between wages and worker productivity unsustainable in equilibrium).

References

Abadie, Alberto, "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 01 2005, *72* (1), 1–19.

Angrist, Joshua D. and Jonathan Guryan, "Does teacher testing raise teacher quality? Evidence from state certification requirements," *Economics of Education Review*, October 2008, *27* (5), 483–503.

_ , Erich Battistin, and Daniela Vuri, "In a Small Moment: Class Size and Moral Hazard in the Mezzogiorno," NBER Working Papers 20173, National Bureau of Economic Research, Inc May 2014.

Behrman, Jere R., Susan W. Parker, Petra E. Todd, and Kenneth I. Wolpin, "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools," *Journal of Political Economy*, 2015, *123* (2), pp. 325–364.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, "Teacher-Student Matching and the Assessment of Teacher Effectiveness," *Journal of Human Resources*, 2006, *41* (4).

del Refugio Guevara, Maria and Laura Elena Gonzalez, "Atraer, formar y retener profesorado de calidad. Reporte sobre la situación de México," Technical Report, OECD 2004.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools," *Journal of Public Economics*, 2015, *123* (0), 92 – 110.

Durante, Ruben, Giovanna Labartino, and Roberto Perotti, "Academic Dynasties: Decentralization, Civic Capital and Familism in Italian Universities," Technical Report November 2014.

Elizondo, Carlos, *Por eso estamos como estamos*, Random House Mondadori, 2011.

Glewwe, Paul W., Eric A. Hanushek, Sarah D. Humpage, and Renato Ravina, "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010," NBER Working Papers 17554, National Bureau of Economic Research, Inc October 2011.

Goldhaber, Dan, "Everyone is Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness?," *Journal of Human Resources*, 2007, *XLII* (4), 765–794.

Hanushek, Eric A. and Steven G. Rivkin, *Teacher Quality*, Vol. 2 of *Handbook of the Economics of Education*, Elsevier, June

Heckman, James J, Hidehiko Ichimura, and Petra E Todd, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, October 1997, *64* (4), 605–54.

Holland, Paul, "Assesing Unusual Agreement Between the Incorrect Answers of Two Examinees Using the K-Index," *Program Statistics Research*, February 1996.

Hoxby, Caroline Minter, "How Teachers' Unions Affect Education Production," *The Quarterly Journal of Economics*, August 1996, *111* (3), 671–718.

Jacob, Brian A. and Lars Lefgren, "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education," *Journal of Labor Economics*, 2008, *26*, 101–136.

_ **and Steven D. Levitt**, "Rotten Apples: An Investigation Of The Prevalence And Predictors Of Teacher Cheating," *The Quarterly Journal of Economics*, August 2003, *118* (3), 843–877.

Kane, Thomas, Jonah Rockoff, and Douglas Staiger, "What does certification tell us about teacher effectiveness?," *Economics of Education Review*, 2008, *27* (6), 615–631.

Kingdon, Geeta and Francis Teal, "Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach," *Journal of Development Economics*, March 2010, *91* (2), 278–288.

Lovenheim, Michael F., "The Effect of Teachers' Unions on Education Production: Evidence from Union Election Certifications in Three Midwestern States," *Journal of Labor Economics*, October 2009, *27* (4), 525–587.

Metzler, Johannes and Ludger Woessmann, "The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation," *Journal of Development Economics*, 2012, *99* (2), 486–496.

Meyer, Bruce D, "Natural and Quasi-experiments in Economics," *Journal of Business & Economic Statistics*, April 1995, *13* (2), 151–61.

Mizala, Alejandra and Hugo Ñopo, "Measuring the Relative Pay of Latin American School Teachers at the turn of the 20th Century," Working Papers 2014-15, Peruvian Economic Association August 2014.

Muralidharan, Karthik and Venkatesh Sundararaman, "Contract Teachers: Experimental Evidence from India," NBER Working Papers 19440, National Bureau of Economic Research, Inc September 2013.

OECD, "Mexico Country Note - Education at a Glance 2014: OECD Indicators," Education at a Glance Country Notes, Organization for Economic Cooperation and Development October 2014.

Oyer, Paul and Scott Schaefer, *Personnel Economics: Hiring and Incentives*, Vol. 4 of *Handbook of Labor Economics*, Elsevier,

Pais, Fundacion Este, "Lo que piensan nuestros maestros: Encuesta National sobre Actitudes y Valores de Maestros y Padres de Familia," *Revista Este Pais*, April 2005.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, 03 2005, *73* (2), 417–458.

Rockoff, Jonah E. and Cecilia Speroni, "Subjective and objective evaluations of teacher effectiveness: Evidence from New York City," *Labour Economics*, 2011, *18* (5), 687–696.

_ , Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger, "Can You Recognize an Effective Teacher When You Recruit One?," *Education Finance and Policy*, January 2011, *6* (1), 43–74.

Rothstein, Jesse, "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *The Quarterly Journal of Economics*, February 2010, *125* (1), 175–214.

Santibanez, Lucrecia, "Reforma Educativa: El papel del SNTE," *Revista Mexicana de Investigacion Educativa*, April-June 2008, *13* (037), 419–443.

_ and Brenda Jarillo, "Conflict and Power: the Teachers'Union and Education Quality in Mexico," *Well-being and Social Policy*, July 2007, *3* (2), 21–40.

Valle, Sonia Del, "Exige CNTE plazas... para venderlas," *Reforma*, June 2015.

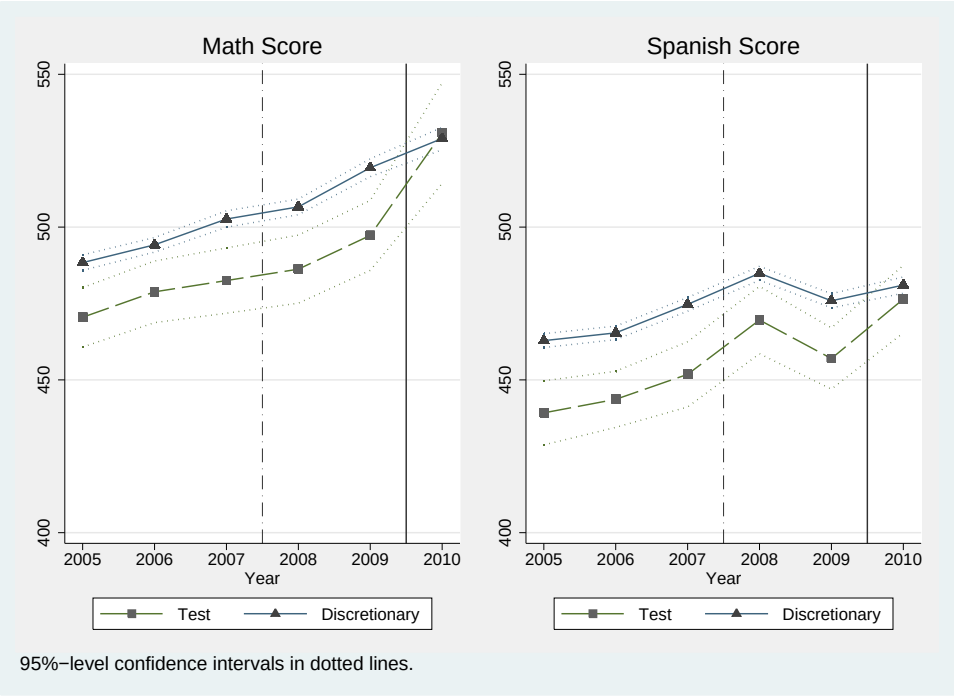
Wayne, Andrew J. and Peter Youngs, "Teacher Characteristics and Student Achievement Gains: A Review," *Review of Educational Research*, 2003, *73* (1), 89–122.

Weissenstein, Michael, "Mexican Senate passes major education reform," *Associated Press*, September 2013.

Zinovyeva, Natalia and Manuel Bagues, "The Role of Connections in Academic Promotions," *American Economic Journal: Applied Economics*, 2015, *7* (2), 264–92.

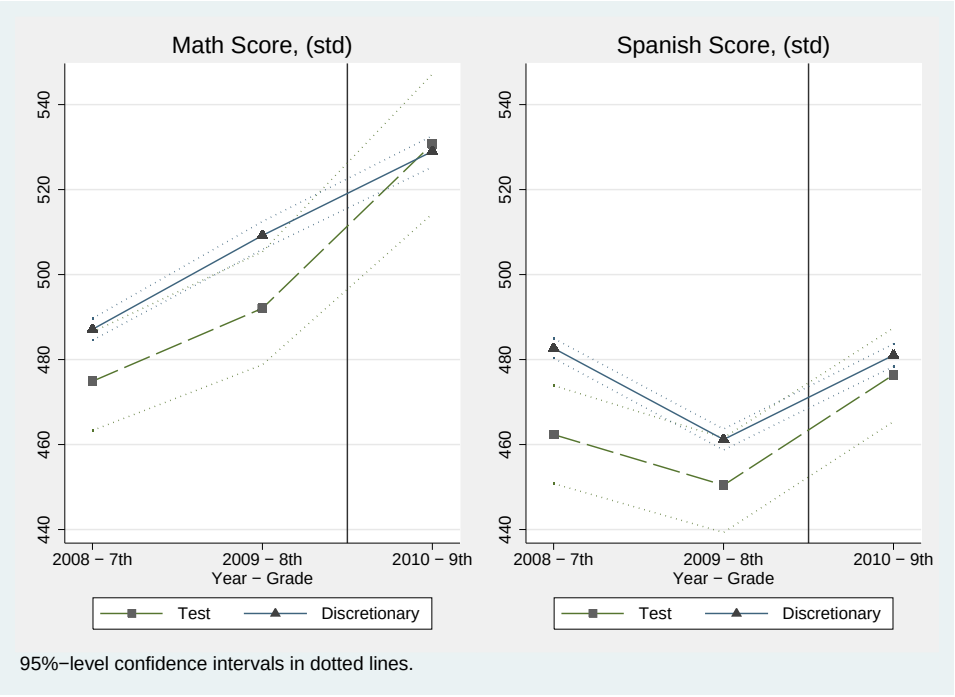
Figures

Figure 1: Evolution of Exam Scores in Schools with New Teachers in 2010



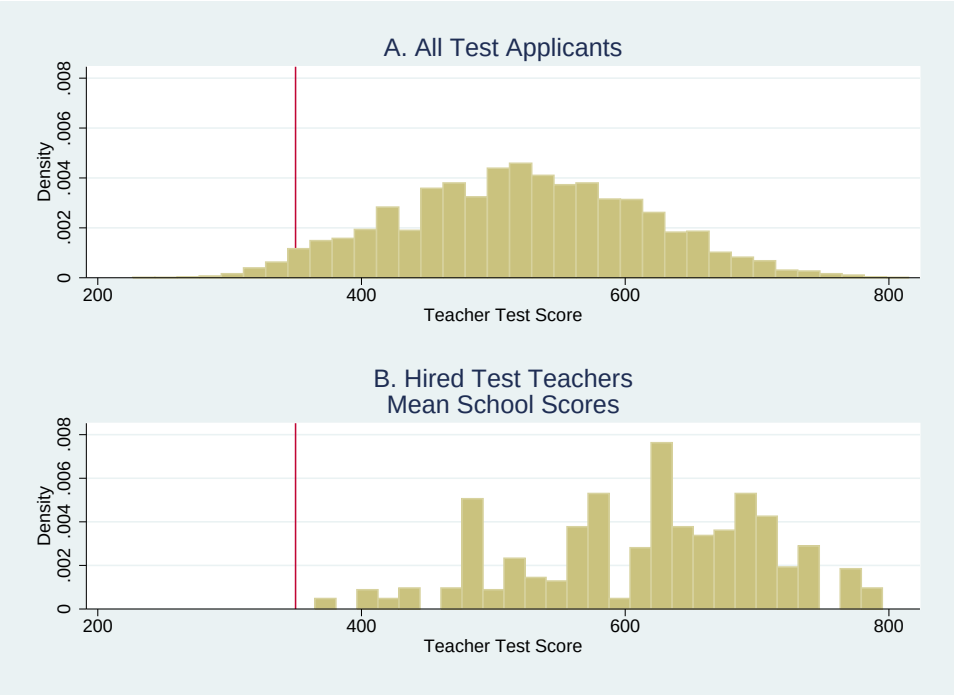
Notes: The graph plots the evolution of yearly means of school (9th grade) scores in the Enlace exam by eventual treatment status. The vertical dotted line indicates the periods before and after test-based hiring was introduced in public schools. The vertical solid line indicates the periods before and after treatment. Sample is composed of Telesecundaria schools which received a new teacher in the year 2010. Enlace scores are standardized at the national level with mean 500 and standard deviation 100. Source: Enlace exam. Enlace results for 2005 corresponds to the academic year 2005-2006 and so on.

Figure 2: Evolution of Exam Scores in Schools with New Teachers in 2010: Panel of Classrooms



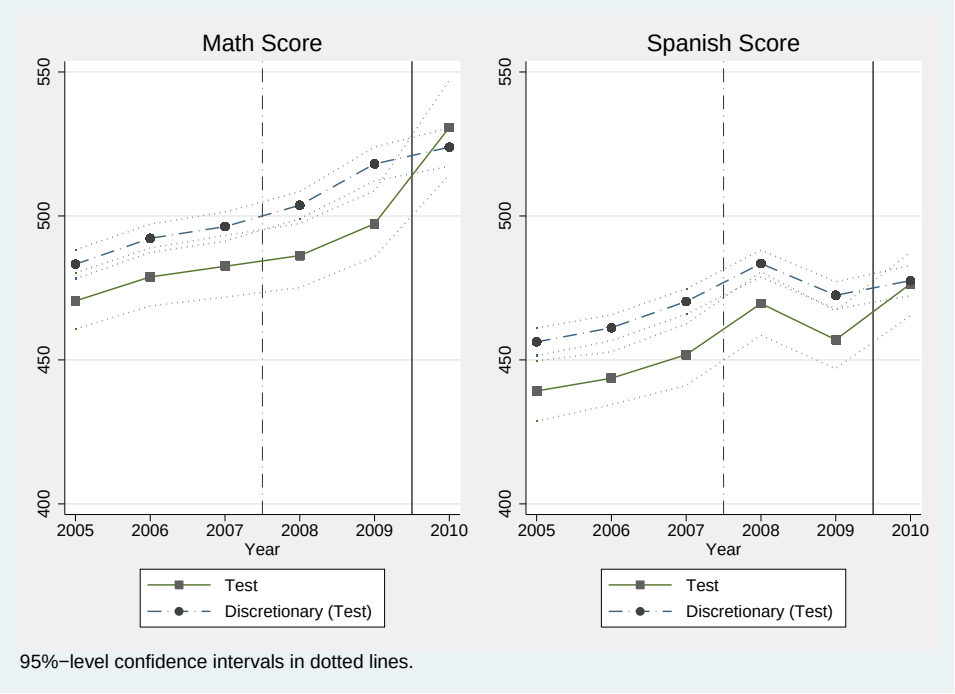
Notes: The graph plots the evolution of yearly means of grade scores in the Enlace exam by eventual treatment status. The vertical solid line indicates the periods before and after treatment. Sample is composed of Telesecundaria schools which received a brand-new teacher in the year 2010. Enlace scores are standardized at the national level with mean 500 and standard deviation 100. Source: Enlace exam. Enlace results for 2008 correspond to the academic year 2008-2009 and so on. Results for 2008 correspond to 7th grade classrooms and so on.

Figure 3: Density of Teacher Exam Scores



Notes: Panel A presents the distribution of teacher test scores for all applicants to Telesecundaria positions in the 15 states in the sample. Panel B presents the school mean of teacher test scores from those applicants hired through the test and allocated to schools in the working sample. The red vertical lines show the (approximate) threshold below which applicants cannot be hired. Scores are standardized at the state level with mean 500 and standard deviation 100. Source: Results from the Concurso Nacional de Asignacion de Plazas Docentes 2010 retrieved on line from the website of the Federal Secretary of Education.

Figure 4: Evolution of Exam Scores in Schools with New Teachers in 2010 Who Applied to the Test-based Hiring



Notes: The graph plots the evolution of yearly means of school (9th grade) scores in the Enlace exam by eventual treatment status. The label Discretionary(Test) refers to schools with discretionary-hired teachers who applied to the test-based hiring (in any year and for any teaching position). The vertical dotted line indicates the periods before and after test-based hiring was introduced in public schools. The vertical solid line indicates the periods before and after treatment. Sample is composed of Telesecundaria schools which received a new teacher in the year 2010. The lines show the evolution of yearly means of school (9th grade) scores in the Enlace exam by eventual treatment status. The label Discretionary refers to school with discretionary-hired teachers who did not participate in the exam-based hiring, while the label Discretionary(Test) refers to schools with discretionary-hired teachers who have participated in the exam-based hiring (in any year and for any teaching position). Enlace scores are standardized at the national level with mean 500 and standard deviation 100. Source: Enlace exam. Enlace results for 2005 corresponds to the academic year 2005-2006 and so on.

Tables

Table 1: Relative Wages of Primary and Junior Secondary Teachers

	Wage (pesos)		Wage gap (ln hourly wage)				
	Monthly	Hourly	Raw		Adjusted		
	Mean	Mean	Beta	SE	Beta	SE	N
Public Prim.-JS Teachers	8,280	70.7					24,599
Private Prim.-JS Teachers	7,110	56.5	-0.2470	0.0207	-0.2087	0.0195	2,487
Other Teachers	8,767	74.7	0.0083	0.0087	-0.0137	0.0084	20,215
Managers	13,285	73.3	-0.0226	0.0101	-0.0310	0.0100	13,405
Others	9,670	53.6	-0.3603	0.0062	-0.3111	0.0061	133,081
Observations	193,787						

Notes: Columns 1 and 2 report mean monthly and hourly wages of college-educated wage earners by occupation categories. Columns 3 and 4 the difference in means (and the standard error) between the base category (in the first row) and the others. Columns 5 and 6 report the coefficients (and standard errors) for the occupation categories in a Mincerian wage equation (the baseline category is in the first row). Source: National Labor Force Survey (ENOE) of the third quarter of 2006 to the second quarter of 2010. The top and bottom 2 percentiles of the wage distibution are trimmed. Sample: Individuals with college education who report positive wage earnings. *** p<0.01, ** p<0.05, * p<0.1

Table 2: Characteristics of Schools with New Teachers in 2010: Means in 2009 (one year before treatment)

VARIABLES	(1) Difference	(2) Test Teacher	(3) Discretionary Teacher
Math Score (std)	-22.14***	497.29	519.42
	(5.81)		
Spanish Score (std)	-18.89***	457.00	475.89
	(4.66)		
Flagged Exams (%)	-0.00	0.05	0.05
	(0.01)		
Final Enrollment (%)	-0.01	0.93	0.94
	(0.01)		
School size	-26.39**	87.86	114.25
	(10.17)		
Class size	-0.66	19.07	19.73
	(0.78)		
Share indigenous students	0.06	0.16	0.10
	(0.03)		
Principal has grad school	-0.12*	0.20	0.32
	(0.05)		
Locality Population	-7584.33	6198.68	13783.01
	(4021.45)		
Hours to state capital	0.72***	2.81	2.09
	(0.11)		
Locality Poverty Rate	0.11***	0.74	0.62
	(0.02)		
Share hhs electricity	-0.05***	0.90	0.95
	(0.01)		
Share hhs sewage	-0.10**	0.64	0.73
	(0.03)		
Observations	1426	95	1331

Notes: All school statistics are for the school year 2009 (one year before treatment). Column 1 reports standard errors, in parenthesis, for a t-test on the equality of means in columns 2 and 3. Enlace scores, suspected cheating and final enrollment correspond to 9th grade results. Enlace scores are standardized at the national level with mean 500 and standard deviation 100. Flagged exams report the share of exams flagged by a cheating detection algorithm run by the Federal Secretary of Education. Final enrollment is the number of Enlace takers over the number of students enrolled at the beginning of the academic year. Sample is composed of Telesecundaria schools which received a new teacher in the year 2010. Source: Enlace, school census and population census 2010. Enlace and school census results for 2009 correspond to the academic year 2009-2010. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Probability of Receiving a Test Teacher (OLS)

VARIABLES	(1) Received Test Teacher in 2010	(2) se
Math Score Lag 1	-1.90e-05	(0.000187)
Spanish Score Lag 1	-7.67e-05	(0.000239)
Math Score Lag 2	-0.000202	(0.000202)
Spanish Score Lag 2	-0.000113	(0.000258)
Students Lag 1	-0.000149*	(7.87e-05)
Class size Lag 1	0.00263*	(0.00139)
Share indigenous students Lag 1	-0.0941***	(0.0334)
Principal has grad school Lag 1	-0.00788	(0.0114)
Locality Poverty Rate	0.108*	(0.0592)
Hours to state capital	0.0299**	(0.0119)
Locality Population	1.88e-07	(1.60e-07)
Share hhs electricity	-0.250**	(0.122)
Share hhs sewage	0.00265	(0.0333)
Observations	1,413	
R-squared	0.161	
State Fixed Effects	Yes	
F statistic Ho Var 1-4=0	1.305	
Prob > F	0.266	

Notes: Enlace scores correspond to 9th grade results and are standardized at the national level with mean 500 and standard deviation 100. Sample is composed of Telesecundaria schools which received a new teacher in the year 2010. Source: Enlace, school census and population census 2010. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Difference-in-Differences: Pre-Treatment Trends in Outcomes

VARIABLES	(1) Final Enrollment (%)	(2) Flagged Exams (%)	(3) Math Score (std)	(4) Spanish Score (std)
2006 X Share Test Teachers	0.0131	0.0544	-2.918	-10.84
	(0.0310)	(0.0331)	(9.854)	(10.11)
2007 X Share Test Teachers	-0.0119	-0.00801	-6.941	-6.335
	(0.0394)	(0.0463)	(13.91)	(15.20)
2008 X Share Test Teachers	0.105***	0.0310	-15.52	-9.891
	(0.0386)	(0.0554)	(13.82)	(13.61)
2009 X Share Test Teachers	-0.00256	0.00203	-2.667	-2.658
	(0.0336)	(0.0368)	(15.12)	(12.87)
Observations	6,952			
R-squared	0.072	0.040	0.127	0.105
Number of id	1,414			
F statistic Ho Var 1-4=0	3.269	0.901	0.594	0.424
Prob>F	0.0111	0.462	0.667	0.791

Notes: Final enrollment (column 1) is the number of Enlace takers over the number of students enrolled at the beginning of the academic year. Flagged exams (column 2) report the share of exams flagged by a cheating detection algorithm run by the Federal Secretary of Education. Enlace scores (columns 3 and 4) are standardized at the national level with mean 500 and standard deviation 100. Regressions include the number of 2010 new teachers in the school, class size, school size, the share of indigenous students, an indicator for principal's attendance of graduate school, a vector of interactions between year and state dummies, and school fixed effects. Results are for 9th grade outcomes. Source: Enlace 2006-2011, school census data 2006-2011 and Registro Maestros 2010-2011. Enlace and school census results for 2005 correspond to the academic year 2005-2006 and so on. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Difference-in-Differences: Results

VARIABLES	(1) Final Enrollment (%)	(2) Flagged Exams (%)	(3) Math Score (std)	(4) Spanish Score (std)
Share Test Teachers	-0.0136	-0.0582**	52.00***	30.66***
	(0.0304)	(0.0225)	(15.97)	(11.21)
Observations	8,365			
R-squared	0.068	0.038	0.163	0.108
Number of id	1,415			
Mean Comparison in 2010	0.937	0.0327	529.0	481.0

Notes: Final enrollment (column 1) is the number of Enlace takers over the number of students enrolled at the beginning of the academic year. Flagged exams (column 2) report the share of exams flagged by a cheating detection algorithm run by the Federal Secretary of Education. Enlace scores (columns 3 and 4) are standardized at the national level with mean 500 and standard deviation 100. Regressions include the number of 2010 new teachers in the school, class size, school size, the share of indigenous students, an indicator for principal's attendance of graduate school, a vector of interactions between year and state dummies, and school fixed effects. Results are for 9th grade outcomes. Source: Enlace 2006-2011, school census data 2006-2011 and Registro Maestros 2010-2011. Enlace and school census results for 2005 correspond to the academic year 2005-2006 and so on. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Difference-in-Differences: Other Outcomes

VARIABLES	(1) Share Indigenous Students	(2) Share Repeaters	(3) Principal has grad school	(4) Class size
Share Test Teachers	0.0393	-0.00262	-0.0489*	-0.0414
	(0.0344)	(0.00687)	(0.0279)	(0.381)
Observations	8,373			
R-squared	0.045	0.030	0.050	0.359
Number of id	1,415			

Notes: Outcomes are measured at the beginning of the academic year. Regressions include the number of 2010 new teachers in the school, school size, a vector of interactions between year and state dummies, and school fixed effects. Source: School census data 2006-2011 and Registro Maestros 2010-2011. School census results for 2005 corresponds to the academic year 2005-2006 and so on. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Difference-in-Differences: Robustness Checks

VARIABLES	(1) Final Enrollment (%)	(2) Flagged Exams (%)	(3) Math Score	(4) Spanish Score
A. School-Specific Linear Time Trends				
Share Test Teachers	-0.0136	-0.0582**	52.04***	30.70***
	(0.0304)	(0.0226)	(15.97)	(11.21)
Observations	8,365			
R-squared	0.064	0.035	0.044	0.062
Number of id	1,415			
B. Binary Treatment (Small Schools)				
2010 Test Teacher	0.00228	-0.0371**	35.81***	19.92***
	(0.0160)	(0.0149)	(10.09)	(6.996)
Observations	3,867			
R-squared	0.078	0.045	0.182	0.114
Number of id	656			
Mean Control in 2010	0.944	0.0330	521.2	475.9
C. Abadie Semiparametric DID Estimator (Small Schools)				
Test Teachers	0.0155	-0.0104	28.03***	16.77**
	(0.0174)	(0.0188)	(9.803)	(7.400)
Observations	468			

Notes: Results are for 9th grade outcomes. Final enrollment (column 1) is the number of Enlace takers over the number of students enrolled at the beginning of the academic year. Flagged exams (column 2) is the share of exams flagged by a cheating detection algorithm run by the Federal Secretary of Education. Enlace scores (columns 3 and 4) are standardized at the national level with mean 500 and standard deviation 100. Panel A: Outcomes are de-trended using school-specific linear time trends. Panels A and B: Regressions include the number of 2010 new teachers in the school, school size, a vector of interactions between year and state dummies, and school fixed effects. Panels B and C: Small schools are those with 3 or less classrooms in 2010. Panel C reports results of Abadie's (2005) semi-parametric difference-in-differences estimator. The propensity score is constructed using a logistic regression in which treatment status is regressed on the covariates listed in Table 2 plus two lags of the change in final enrollment and exam cheating. Third-order polynomials are included for all continuos variables. Estimation of results is restricted to observations in the common support of the propensity score. Source: Enlace 2006-2011, school census data 2006-2011 and Registro Maestros 2010-2011. Enlace and school census results for 2005 corresponds to the academic year 2005-2006 and so on. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Difference-in-Differences: Panel of Classrooms

VARIABLES	(1) Final Enrollment (%)	(2) Share Flagged Exams (%)	(3) Math Score (std)	(4) Spanish Score (std)
A. Pre-Treatment Outcomes				
Share Test Teachers	-0.000409	0.0150	-17.43	4.095
	(0.0341)	(0.0260)	(13.01)	(9.872)
Observations	2,799			
R-squared	0.057	0.092	0.152	0.248
Number of id	1,414			
Mean Control in 2009	0.930	0.0454	509.2	461.2
B. Results				
Share Test Teachers	0.0203	-0.0468***	42.68***	20.76**
	(0.0345)	(0.0174)	(14.77)	(10.06)
Observations	4,212			
R-squared	0.043	0.056	0.212	0.154
Number of id	1,415			
Mean Control in 2010	0.937	0.0327	529.0	481.0

Notes: Regressions include the number of 2010 new teachers in the school, class size, school size, the share of indigenous students, an indicator for principal's attendance of graduate school, a vector of interactions between year and state dummies, and school fixed effects. Final enrollment (column 1) is the number of Enlace takers over the number of students enrolled at the beginning of the academic year. Flagged exams (column 2) report the share of exams flagged by a cheating detection algorithm run by the Federal Secretary of Education. Enlace scores (columns 3 and 4) are standardized at the national level with mean 500 and standard deviation 100. Source: Enlace 2009-2011, school census data 2009-2011 and Registro Maestros 2010-2011. Enlace and school census results for 2008 corresponds to the academic year 2008-2009 and so on. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Difference-in-Differences: Results – 25th and 75th Percentiles

VARIABLES	(1) Math Score pct 25	(2) Math Score pct 75	(3) Spanish Score pct 25	(4) Spanish Score pct 75
Share Test Teachers	54.98***	51.00***	29.47**	31.29***
	(16.92)	(19.35)	(13.33)	(11.48)
Observations	8,365			
R-squared	0.131	0.171	0.078	0.099
Number of id	1,415			
Mean Control	466.8	589.6	422.7	537.2

Notes: Outcomes in columns 1 and 2 (3 and 4) are the 25th and 75th percentiles of the school Enlace score distribution in mathematics (Spanish). Enlace scores are standardized at the national level with mean 500 and standard deviation 100. Regressions include the number of 2010 new teachers in the school, class size, school size, the share of indigenous students, an indicator for principal's attendance of graduate school, a vector of interactions between year and state dummies, and school fixed effects. Results are for 9th grade outcomes. Source: Enlace 2006-2011, school census data 2006-2011 and Registro Maestros 2010-2011. Enlace and school census results for 2005 corresponds to the academic year 2005-2006 and so on. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Teachers Test Scores and Student Achievement

VARIABLES	(1) Math Score (std)	(2) Spanish Score (std)	(3) Math Score (std)	(4) Spanish Score (std)
Teacher Exam Score	-0.00500	0.0381	0.0266	-0.0184
	(0.138)	(0.102)	(0.128)	(0.0799)
Observations	545	545	278	278
R-squared	0.339	0.270	0.410	0.286
Number of id	94	94	94	94
Panel	Schools	Schools	Classrooms	Classrooms

Notes: Regressions include the number of 2010 new teachers in the school, class size, school size, the share of indigenous students, an indicator for principal's attendance of graduate school, a vector of interactions between year and state dummies, and school fixed effects. Enlace scores are standardized at the national level with mean 500 and standard deviation 100. Source: Enlace 2006-2011, school census data 2006-2011 and Registro Maestros 2010-2011. Enlace and school census results for 2005 corresponds to the academic year 2005-2006 and so on. Standard errors in parentheses are clustered at the school level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 11: Difference-in-Differences: Teachers Who Applied to the Test-based Hiring

VARIABLES	(1) Math Score (std)	(2) Spanish Score (std)	(3) Math Score (std)	(4) Spanish Score (std)
A. Pre-Treatment Outcomes				
2006 X Share Test Teachers	-3.708	-12.69		
	(10.10)	(10.64)		
2007 X Share Test Teachers	-10.78	-10.89		
	(14.30)	(15.65)		
2008 X Share Test Teachers	-19.31	-9.067		
	(14.04)	(13.72)		
2009 X Share Test Teachers	-2.734	-6.573		
	(14.93)	(13.19)		
Observations	4,793			
R-squared	0.134	0.116		
Number of id	972			
F statistic Ho Var 1-4=0	0.895	0.379		
Prob > F	0.466	0.824		
B. Results				
Share Test Teachers	55.12***	36.14***	56.42***	39.30***
	(16.65)	(12.09)	(16.70)	(12.17)
2010 New Teachers' College GPA			-0.0166	-0.0430*
			(0.0347)	(0.0261)
2010 New Teachers' Age			0.427	0.634
			(0.636)	(0.457)
Observations	2,408			
R-squared	0.200	0.150	0.200	0.152
Number of id	412			
Mean Control in 2010	523.9	477.5	523.9	477.5

Notes: Regressions include the number of 2010 new teachers in the school, class size, school size, the share of indigenous students, an indicator for principal's attendance of graduate school, a vector of interactions between year and state dummies, and school fixed effects. Final enrollment (column 1) is the number of Enlace takers over the number of students enrolled at the beginning of the academic year. Flagged exams (column 2) report the share of exams flagged by a cheating detection algorithm run by the Federal Secretary of Education. Enlace scores (columns 3 and 4) are standardized at the national level with mean 500 and standard deviation 100. Source: Enlace 2006-2011, school census data 2006-2011 and Registro Maestros 2010-2011. Enlace and school census results for 2005 corresponds to the academic year 2005-2006 and so on. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1