Rules Rather Than Discretion: Teacher Hiring and Rent Extraction

Ricardo Estrada

Paris School of Economics*

May 20, 2013

Abstract

There is mounting research evidence on both the importance of teacher quality in the production of learning and on the difficulty of identifying who actually is (can be) a good teacher. The participation of current teachers in the selection of new teachers may ease this informational problem e.g. if the former have superior information about teacher quality. However, a simple principal-agent model can show that asymmetric information between administrators and teachers may lead to an agency problem with rent extraction if teachers have an objective function with different arguments from teacher quality. In this study, I use a recent policy reform in Mexico to evaluate the effect on student outcomes of receiving a teacher hired through a standardized test versus one hired in a discretionary process with strong involvement from the teachers' union. My difference-in-differences results indicate that the allocation of test-hired teachers reduces exam cheating in junior-secondary schools and -in environments where cheating is rare- increases student achievement. I also find that joint committees of state officials and union representatives allocate the discretionary-hired teachers to schools in more "desirable" localities. Taken together, these results suggest the existence of an agency problem with potential rent-extraction that the use of a hiring rule can mitigate.

JEL: D82, I10, J23.

Keywords: teacher quality, rent extraction, teacher labor markets, teachers' unions, student achievement, exam cheating, Mexico.

*48 boulevard Jourdan, 75014 Paris, restrada@pse.ens.fr. I am grateful to Marc Gurgand, Jérémie Gignoux, François Bourguignon, Fabian Lange, Francisco Ferreira, Paul Glewwe, Jorge Santibanez, Kenneth Houngbedji, Alejandro del Valle and participants at seminars at the Paris School of Economics and the Lisbon ISEG workshop for insightful comments. I thank the staff at Mexicanos Primero for kindly providing me with access to data.

1 Introduction

There is mounting research evidence on both the importance of teacher quality in the educational production function and on the difficulty of identifying at the time of hire who actually is (can be) a good teacher. The problem for school administrators (and parents) is that the characteristics they can typically observe at the time of hire, like experience and training certificates, are poor predictors of teacher quality -or irrelevant for recent graduates (see, for example, *Hanushek and Rivkin* (2006)).

The participation of current teachers in the selection of new teachers may ease this informational problem if the former have either superior information or higher ability to identify teacher quality using broader criteria.[1] However, asymmetric information between administrators and current teachers may lead to an agency problem if current teachers' optimize an objective function with different arguments from teacher quality. In this case, administrators would face a trade-off between hiring using rules that are second-best predictors of quality, but hard to manipulate, or giving discretionary powers to an agent endowed with better information, but with a distinct objective function.[2]

In this paper, I use a recent policy reform in Mexico to evaluate the effect on student outcomes of receiving a new teacher hired through a standardized test versus one hired in a discretionary process with strong participation from the teachers' union. The recruitment of teachers for public primary and junior-secondary schools in Mexico is centralized at the state level and is not directly associated to the filling-in of specific vacancies in schools. Prior to the reform, state officials would select almost the totality of new teachers required in the state in a discretionary process with strong involvement from the teachers' union. With the reform, a share of the new teachers is selected on the basis of a standardized test given for this purpose.

The reform did not change the mechanism to allocate teachers to schools. State-wide committees jointly chaired by state officials and representatives from the teachers' union are in charge at the beginning of every academic year to fill in from the stock of old and newly hired teachers the vacancies opened in schools by retirement, between-school transfers and the expansion of teaching positions. Joint committees allocate current teachers to available schools based on teachers' applications for specific school positions and a set of pre-defined criteria. New teachers are allocated to available schools after current teachers' choices. A priori, whether test-hired and discretionary-hired teachers are allocated to schools with different characteristics is an empirical question.

I investigate the allocation of test and discretionary teachers to junior secondary schools using panel

[1] For example, Rockoff, Jacob, Kane and Staiger (2011) find that combining a broad set of measures of teacher characteristics is informative about the effectiveness of new math teachers in New York City, while the same measures have little predictive power as independent factors.

[2] *Hoxby* (1996)discusses a theoretical model in which teachers demand unions to influence the educational production function either because: 1) though they have the same objective function than administrators and parents (maximization of student achievement), teachers have better information about input efficiency or internalize externalities than the others neglect; alternatively 2) teachers may have a different objective function than administrators and parents, and hence a desire to set the school inputs that maximize their own objectives.

data with five yearly observations per school before treatment.[3] I do not find pre-treatment differential trends between schools where test and discretionary teachers are assigned when I run a school fixed-effects model in which I regress (school-level) student outcomes in a standardized exam on a vector of year dummies interacted with treatment status and a set of time-variant school inputs and state effects. I believe this evidence firmly supports the plausibility of the "parallel trend" assumption necessary for the identification of a causal treatment effect in a difference-in-differences model.

In contrast, I find in a cross-sectional OLS regression that discretionary teachers are more likely to be assigned to schools located in "more desirable" localities -less poor, closer to state capitals and with larger penetration of public services. This result is consistent with a model in which 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.

My difference-in-differences estimates indicate that students benefit from having test teachers. Students in schools that receive test teachers are less likely, by 3.4 percentage points, to cheat in an end of junior high school standardized exam than students in schools that receive discretionary teachers. The treatment effect amounts to a 50 percent reduction of the baseline rate. In a similar fashion, going from not having test teachers in a school to having only test teachers is associated with a reduction of 6.99 percentage points in the share of students detected for exam copying, an effect that is statistically significant at the one-percent level.

The Federal Ministry 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.

The result is robust when I use a non-parametric DID estimator (Abadie, 2005) which allows me to deal with concerns for potential bias in simple DID estimation introduced by heterogeneous treatment effects along characteristics for which there is an unbalanced distribution between the two groups of schools (Heckman, Ichimura and Todd 1998 and Meyer 1995).

When I focus my estimation in a set of schools in which exam copying is rare, I find that the allocation of a test teacher has a positive and sizable effect on student achievement. Moving from no test teachers in a school to only test teachers is associated with an increase in .663 standard deviations in the school's Mathematics test score (.788 for the Spanish score), a result that is statistically significant at the five-percent level (one-percent level for the Spanish score case).

Summing up, my findings indicate that education officials hire on average teachers of less quality when they follow a discretionary process with participation from the teachers' union. Then, joint committees of state officials and union representatives allocate these teachers to schools in more "desirable" localities.

[3] I restrict my analysis to a type of rural schools called Telesecundarias for data limitations.

Taken together, these results suggest the existence of an agency problem with potential rent-extraction. State officials are able to reduce the extent of this problem by using a hard-to-manipulate rule to hire new teachers -a ranking based on a teacher test. These findings are particularly relevant for environments with weak institutions, where the lack of accountability (on actual teacher performance, for example) may exacerbate the incentives for rent-seeking behavior arising because of imperfect information about worker productivity.

2 Hiring as an agency problem

2.1 Theoretical Model

I write down a very simple principal-agent model with the purpose of illustrating the structural tension that may arise when a principal evaluates engaging an agent to perform the task of hiring a worker in an economy with asymmetric information about worker productivity.

Consider a principal p who engages agent a to hire a teacher i from a continuum of potential applicants with teacher quality uniformly distributed on a segment represented by [0,1] and expressed in (normalized) monetary terms. Suppose the principal does not observe teacher quality, while the agent observes the whole distribution of teacher quality. (Potential) teachers know their own quality. The process to fill the vacancy is straightforward: 1) the principal receives an exogenous endowment ($w > 0$) and posts a vacancy with wage offer w, 2) would-be teachers with $q_i \leq w$ apply to the vacancy and 3) the agent selects applicant i of quality q_i from the segment [0,w].

Suppose the principal's utility function is $V_p = q_i$, where q_i is the quality of hire i and $\frac{\partial V}{\partial q} > 0$. Now, the agent may derive positive utility from teacher quality and from a rent (r_i) that the agent can extract if the posted wage is larger than the hire's quality. Say $r_i = w - q_i$ if the the agent does not share the rent with the hire, otherwise the agent gets $r_i = \alpha (w - q_i)$ and the hire gets $(1 - \alpha)(w - q_i)$. Assuming Cobb-Douglas preferences and $\alpha = 1$, we can write the agent's optimization program as:

$$\max_{q,r} U_a = q_i^{\gamma} r_i^{(1-\gamma)} \qquad s.t. \qquad w = q_i + r_i \tag{1}$$

Where γ is an exogenous preference parameter that captures the agent's taste for teacher quality.

Hoxby (1996) discusses a theoretical framework in which teachers demand unions to influence the educational production function either because 1) in a "efficiency-enhancing" model of participation, teachers share the objective function of administrators 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.

It is clear that if $\gamma \in \{0, 1\}$, the described principal-agent model fits Hoxby's framework. Her efficiency-enhancing model of unions' involvement in education arises when $\gamma = 1$, while the rent-seeking model corresponds to $\gamma = 0$. If $\gamma = 1$, the principal and the agent have the same utility function ($V_p = U_a$) and the solution to the agent's optimization problem ($q_i^* = w$) also maximizes the principal's utility given the wage constrain. In contrast, if $\gamma = 0$ the optimal solution for the agent is $q_i^* = 0$, which is suboptimal for the principal.

In this setting, the agency problem depends on what is the agent's type: efficiency-enhancer or rent-seeker. Note that if the principal knows that the agent is a rent-seeker ($\beta = 0$), she is better off by replacing the agent for a very basic selection rule: picking an applicant at random (as $E\left[q_i | 0 \leq q \leq w\right] = \frac{w}{2} > 0$). This is a "second-best" selection rule with respect to the outcome with an fully-informed and preference-aligned agent.[4]

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 at least partially teacher quality ex-post. A simplistic way to bring a dynamic flavor into this setting is to allow γ to capture both the agent's intrinsic 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 γ. If γ is continuous, the interior solution to equation 1 is $q_i^* = \gamma w$. Assuming again that the value of γ is public information, the payoff of using the agent versus the selection rule (γw vs. $\frac{w}{2}$) depends, on the one hand, on the agent's intrinsic taste for teacher quality and internalization of the degree in which the principal observes teacher quality ex-post, and on the other hand, in the predictive power (expected value) of the selection rule available to the principal.

There are reasons to expect that administrators are better equipped to identify teacher quality ex-post than ex-ante. For example, *Jacob and Lefgren* (2008) find that school principals do well at identifying teachers in the extremes of the quality distribution. More broadly, the growing demand for the identification of good teachers has pushed the development of a whole methodology devoted to asses teacher quality in terms of value-added to students' test scores. Although great progress has been done in this field, specially when teacher quality is measured at some aggregated level, *Rothstein* (2010) shows that conventional measures of individual teacher's value added depend on restrictive conditions, require rich datasets and, hence, are still unlikely to be a reliable input for personnel decisions in many settings.

Given the technological complexity associated to the identification of individual teacher quality, even with ex-post measures, one could expect the described agency problem to be more relevant in environments with weak institutions. In other words, where low institutional capacity or political economy dynamics weaken either the evaluation of teacher performance (worker productivity) or the capacity of

[4] In this simple framework, the agent cannot credible commit to select a candidate $q^* > \frac{w}{2}$ -and induce his engagement by the principal- as once engaged, the agent has the incentive to select $q_i^* = 0$.

the principal to punish the agent for selecting low quality candidates.[5]

2.2 Empirical Approximation

Following Hoxby (1996), I derive an empirical specification to investigate if the discretionary hiring of new teachers follows an "efficiency-enhancing" or "rent-seeking" model of participation in education. Suppose you observe a panel of schools which receive a new teacher at time $t = 1$ hired either through a selection rule or a discretionary process, and you write a model like:

$$y_{st} = \beta T_{st} + \theta X_{st} + \tau_t + \alpha_s + v_{it} \tag{2}$$

Where y_{st} is the average of student achievement at school s at time t, T_{st} is a indicator that equals 1 if school s received a teacher hired using a selection rule at $t = 1$ and 0 otherwise, β is the parameter of interest, X_{st} is a vector of time-variant school inputs, τ_t is a vector of year effects, α_s is a vector of school specific intercepts and v_{st} is a disturbance term.

This is the typical framework for a difference-in differences estimation with panel fixed effects. Under the parallel trend assumption, β is an unbiased estimate of the average treatment effect on student achievement at the school level of receiving a teacher hired using a selection rule versus one hired in a discretionary process.

We can expect $\beta > 0$ if the agent behaves as a rent-seeker and $\beta < 0$ if the agent behaves as an efficiency-enhancer. In this model, β would capture both the agent's intrinsic taste for teacher quality and the internalization of any dynamic incentive related to the principal's capacity to observe teacher quality ex-post and punish the agent for hiring low-quality teachers.

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 classrooms in a school. In the system under study, teachers actually teach only one classroom per school. So, the classroom might be a more natural unit to conceptualize the influence of a teacher. I cannot directly link teachers to classrooms in the data though. Beyond this limitation, the matching between teachers and students and potential within-school externalities can make the identification of a causal treatment effect at the classroom level more restrictive.

In sections 5 and 6.1, I describe and empirically investigate the allocation of teachers to schools and find strong support for the identification assumption behind the estimation of a causal treatment 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 about the assignment of students to teachers in which observational studies rely to identify teacher causal effects.[5]

[5]Although the Rothstein's critic is focused in the estimation of individual teacher effects which require more restrictive

The focus in the classroom level makes also easier to neglect within-school externalities associated to teacher quality. A higher quality teacher could for example free up other school resources -like principal's time- for the benefit of students in other classrooms. Also, higher quality teachers might have a direct effect on students in other classrooms through direct interactions.

Summing up, in the model outlined in equation 2, β is a parameter informative about the average total policy effect of allocating a rule-selected teacher -instead of a discretionary-selected teacher- at the school level. On the one hand, the estimation of a treatment effect at the school level rather than at the classroom level makes identification more compelling. On the other hand, as common sense suggests that the main effect of teachers is concentrated on the students they actually teach, β represents likely an under-estimation of the average difference in quality among the teachers selected under the two discussed mechanisms.

3 A Policy Reform to Teacher Hiring in Mexico

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 states with the bulk of funding.[6] Teacher hiring, allocation to schools and promotions are centralized at the state level. State Ministries of Education are in charge of teacher hiring, which is not related to fill specific school positions. Joint committees of state officials and teachers' union representatives are responsible for the allocation of teachers to schools.

Starting in 2008, the Federal government championed a reform to teacher recruitment for public schools in the country, introducing standardized testing as a mechanism to hire teachers.

3.1 Teacher Hiring Prior to 2008

State Ministries of Education are responsible for hiring the new teachers required to fill-in the vacant positions in the school system. From an aggregated point of view, vacancies equal the expansion of the stock of state teachers plus outflows from this stock -to retirement or other occupations, for example.

It is important to have in mind that though hiring targets quotas of different teacher types (e.g. Primary school teacher, Mathematics JHS teacher etc.), it is in principle not related to specific vacancies at schools. The allocation of teachers to schools is defined later by a joint committee of state officials and teachers' union representatives in a process that I describe in detail below.

Brand-new teachers in public elementary schools must have university-level studies, though not necessarily a degree at the moment of hiring. Teaching in junior high school education is not restricted to

assumptions than the estimation of an aggregated effect.

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

graduates from teacher schools. For the sake of clarity, it is important to have in mind that when I use the term new teacher I refer to a person entering into a teaching position in the public education system. This definition excludes incumbent teachers that are transferred from one public school to another.

The bulk of state vacancies is generated by outflows from the stock of teachers. There is a wide consensus that the State Ministries of Education rely heavily on the teachers' union to hire for these positions (*Santibanez*, 2008). The teachers' union (SNTE by its Spanish acronym) is a national organization with 52 regional sections. Both affiliation and payment of fees to the union is mandatory for all teachers in public elementary schools.

For starters, the union plays an important role in disseminating information among would-be teachers as, before the reform, state governments would rarely publicize the number and type of available vacancies. Also, it is common practice that retiring teachers propose via the union one would-be teacher with priority for hiring. This is probably the main mechanism used in practice to select new teachers. Under a broader law that regulates the labor relations of public employees, the teachers' union is entitled to directly select for hiring a number of new teachers equivalent to the 50% of the yearly expansion in the stock of teachers.

In some cases, State Ministries of Education hire a fixed share of the new graduates from the public teacher schools in the state, using generally the other 50% of the new payroll positions or temporary contracts.

State officials and the teachers' union have been subject to criticism for neglecting teacher quality when hiring. Strong, but not isolated, denounces include the selling of teaching positions and the practice of teachers going into retirement to bequeath their position to a relative. A national survey among elementary teachers found that one-third of interviewees thought that selling of teaching positions was a frequent practice (*Santibanez*, 2008). When the 2008 reform was announced, union leaders in at least two states publicly declared their opposition to the examination because it would take out the union members' right to bequest their position (*Elizondo*, 2011).

3.2 Test-based Hiring

In 2008, the Federal government presented (under the umbrella of a broader agreement with the Teachers' Union) a plan to open to competitive examination all vacant teaching positions in public primary and junior secondary education in the country. The mechanics of the new examination is the following.

Competition is open to candidates willing to enter the teaching profession in public schools and current teachers with temporary contracts. There are hiring quotas for each group. In this paper, I focus only in the recruitment of brand-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 JHS teacher etc.). The standardized exam is designed to measure cognitive skills, knowledge of the teaching subject and mastery of teaching methods.

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 to specific schools. Some type of teaching positions are restricted to graduates of teacher schools or from specific college majors.

The reform met with strong opposition from state officials and local union leaders. In a compromising result, only new payroll positions funded by the Federal Government were filled through the test-based recruitment initially, though it was expected that progressively more vacants were opened to test-based hiring. The Teachers' Union agreed to cede its selection entitlement over the 50% of the federally-funded new payroll positions. Almost all states use the test-based recruitment to fill some of their vacancies since 2008 (30 of 32, including the Federal District). According to figures from the Federal Ministry of Education, from the 22,546 full-time vacancies opened to test hiring in the 2010 school year, 34% corresponded to new positions and the rest to existing payroll positions. There is not public information about the total number of new teachers hired through discretionary recruitment. They could amount to around 82% of all teachers hired according to my estimates.[7]

Hence, every year states select new teachers through both test-based and discretionary recruitment. The reform did not change the mechanism to allocate teachers to schools.

3.3 Telesecundaria Schools

As I explain below, I only observe the link between teachers and students at the school level. Hence, I focus my empirical analysis in Telesecundaria schools, a system of public junior secondary education (Grades 7 to 9). Telesecundarias are small schools catered to small communities. The small school size should increase the likelihood that I find a statistical significant teacher effect at the school level.

The typical Telesecundaria school in my sample has 69 students divided in 3 classrooms and is located in a locality with 817 inhabitants -all are median values. This educational system is widespread, though. According to figures from the Federal Ministry of Education, around 1.26 million students attended 18,000 Telesecundaria schools in the school year 2010, which amounts to 20.6% of student enrollment in junior secondary education in the country.

Telesecundaria students tend to be more rural, poorer and face in general more disadvantaged conditions that 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 69%, while the national poverty rate was 46% -according to the National Council for the Evaluation of Social Policy (CONEVAL).

[7] In the Telesecundaria System.

Contrary to general junior high schools, Telesecundarias have one teacher per classroom -as opposed to one teacher per topic. Instead of specialist teachers, Telesecundarias rely heavily on IT teaching support. The television programs that the Federal Ministry of Education produces specifically for this school system fill approximately 2 of the 6 hours of the school day. Hence, it seems reasonable to assume that the effect of teacher quality in Telesecundarias is lower than in educational systems in which teachers play a larger role in the learning process.

4 Data

4.1 Enlace Exam

I use the results in a national standardized test (Enlace) that students take at the end of the academic year 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.[8]

Test scores for grades 7 and 8 are only available since the school year 2008, as only 9 graders would take the exam in the 2005-2007 period. The grade 9 exam assessed materials of grades 7 to 9 before 2007, while after this year is focus in grade 9 materials. The test measures learning in Mathematics, Spanish and a rotating subject every year. I use the first two subjects for the analysis. I only use the results from grade 9 in my main estimations because the larger panel dimension.

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 Ministry of Education delivers monetary bonuses to teachers of classrooms and in schools in the top 15% in the -respective- score distribution; and to teachers in schools in the top 15% in the score gains distribution (gains with respect to the previous two years). Schools are classified by state into categories defined by locality (urban/rural and with high/low marginalization) and school (general/technical/telesecundaria/etc.) characteristics. A teacher can receive a bonus ranging from $2,000 up to $20,000 pesos (around USD PPP 260 and 2,600, respectively).

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

4.1.1 Detection of Cheating

Students take the Enlace exam in two school days towards the end of the academic year. Each State Ministry of Education allocates one exam coordinator per school to overview along the school principal the implementation of the test. The school principal selects one teacher per classroom to monitor the

[8]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.

students during the application of the exam. It is forbidden that teachers monitor the classrooms they teach.[9] 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 the exam responses by a third-party.

The Federal Ministry of Education runs a software to detect test cheating using two statistical tools commonly used for this purpose, the K-Index and the Error Similarity Analysis (ESA) Index.[10] 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 in 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 more informative of copying.

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 application of 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 do not have incorrect responses. E.g. if some one gives one string of correct answers to the whole students in a classroom. Moreover, both methods are unlikely to be informative about other forms of cheating that may involve students (like use of cheat sheets and impersonation) or teachers (e.g. giving students extra-time or teaching to the test).

There are not sanctions to either principals, teachers or students associated to suspected cheating. The individual exams that are flagged as suspicious of cheating are not taking into account though for the estimation of the school score that is reported in the official results. The Federal Ministry of Education delivers to the State Ministries a report with the list of the schools in which a high prevalence of cheating is detected.

4.2 School and Locality Characteristics

I use the census of schools carried out by the Ministry of Education (Formato 911) to obtain yearly 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

[9] 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.

[10] 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 both methods can be found, respectively, at *Holland* (1996) and Belleza and Belleza (REF).

Social Policy about the localities' poverty rate. I obtain from Google Maps the estimated distance by car from the schools' localities to the State capital.[11]

4.3 Census of Teachers

I benefit from extensive data of school personnel compiled due to 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 Ministry 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 though information about hiring, education profile or assigned classrooms.

I do not observe directly in the data who are the teachers hired since 2008, when the test-based examination was implemented, and how these teachers were recruited. However, I can use the 2009 and 2010 censuses to identify the 2010 cohort of new teachers and match these to the available results of the test-based hiring. Hence, I focus my analysis here and after in the (24) states that opened vacancies for the Telesecundaria system in the 2010 test-based hiring.[12]

I identify the 2010 cohort of new teachers by first comparing the 2010-2nd-quarter census of Telesecundaria's personnel (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 Telesecundaria System's brand-new personnel in the 2010 school year. I drop observations from 7 states that report relatively few personnel in the 2nd quarter of 2010 and hence have a high, an likely unreliable, ratio of new/total personnel (larger than 20%) in the 2010 school year -the mean ratio in the remaining states is 6.7%.[13]

The SEP dataset includes a module with the list of the 2009 and 2010 test-selected applicants (626 and 550 teachers, respectively, in the 24 states). I merge this module to the main dataset using the national population number.[14] I am able to merge 68% (427 observations) of the 2009 and 72% (395 observations) of the 2010 test teachers to specific schools. I also find 12 (.8%) 2010 test teachers in supervision offices. The dataset includes a State Education Ministries' report on the candidates' hiring status. 39% of the the non-matched 2010 test teachers are declared to be in the waiting list for allocation to a school, 18% did not fulfill all the administrative requirements to be hired, 10% did not accept the assigned school and

[11] Using the Stata command traveltime.

[12] 2 states (Michoacan and Oaxaca) do not participate at all in the test-based examination and 6 states did not open to test-based competition any vacancy at the Telesecundaria system in 2010 (Baja California Sur, Colima, Nayarit, Queretaro, Sonora and Zacatecas).

[13] These states are Baja California, Guanajuato, Guerrero, Tabasco, Tlaxcala, Veracruz and Yucatan.

[14] 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.

there is no status information for the remaining 33%.[15]

I find that 10% 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. In the extreme, 15 of the 16 test teachers hired in the state of Nuevo Leon fall in this case. This evidence suggests that some incumbent teachers -maybe hired under temporary contracts- where allowed to participate in the examination for brand-new teachers. I drop out 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.

The database is inclusive of teachers, administrative staff and principals. I identify as teachers all observations where at least in one quarterly database I observe a synonymous of the word "Teacher" or "Hours Telesecundaria" in the two variables with information about the post description .[16]

Overall, I have a database with 1,869 new teachers (19% test-selected) distributed in 1,661 schools in 15 states in the 2010 school year. In addition, I have information on the 415 test teachers hired in 2009 for whom I record the schools where they were allocated in 2009 and 2010 . I collapse then the dataset at the school-year level and merge it to the panel with school results and characteristics.

I obtained from the Federal Ministry of Education a list with the schools where the 2008 test teachers were initially assigned -the file does not have the teachers' population or tax identification number. I add this information to the panel of schools.

After merging, I have a panel with 1,638 schools in 15 states. The size of the database reduces to 1,246 schools in 13 states when I restrict to schools which have never received a 2009 or 2008 test teacher, which did not 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. 12.5% of these schools received a test teacher in the school year 2010.[17]

5 Allocation of Teachers to Schools

In each state, a joint committee of state officials and union representatives is in charge of allocating teachers to schools. Joint committees operate under state-level regulations heavily based on a 1973 agreement between the Federal Ministry of Education (SEP) and the Teachers' Union (SNTE). The 1973 SEP-SNTE agreement stipulates that all vacant positions at schools should be subject to competition among teachers currently employed in the public system. Committees must evaluate candidates according to their certifications, tenure, ability and discipline. Joint committees' decisions are mandatory for ministry and school administrators. The allocation of teachers to schools depends hence on teachers' preferences and committees' assessments of applicants' relative merit.

Broadly, the process works as follows: First, the joint committee announces to current teachers the

[15] I drop out all observations from the Federal District as there is no match among 2010 test-hired teachers.

[16] The actual keywords that I use are: Maestro, Mtro, Profesor, Docente, Horas Telesecundaria and H.S.M. I also use three payroll codes which I know from the data that are associated to teaching positions.

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

list of schools in the state with available positions due to retirements, expansion of the school staff, etc. Second, interested teachers apply to specific schools. Third, the committee awards these positions. Afterwards, a process known as the *corrimiento* takes place. The school positions opened due to the between-school transfers done in the first-stage are now posted for applications. The process is repeated until no incumbent teacher is interested in the available school positions. Then, new teachers are assigned to these schools.

Joint committees do not have to follow the same criteria to allocate newly-hired teachers to schools with teaching vacancies and enjoy more discretionary power in this process. Hence, a priori, whether test and discretionary teachers are allocated to schools with different characteristics is an empirical question.

Table 1 shows descriptive statistics for the set of Telesecundaria schools that received either test (treatment) or discretionary (control) teachers hired in the 2010 school year.[18] It is clear that schools that receive test and discretionary teachers are different, notably in, but not restricted to the characteristics of the localities where they are located. The localities where the schools with test teachers are located are smaller, poorer, further away from the state capital and have less penetration of public services. For example, with respect to discretionary teachers, test teachers are allocated on average to schools located almost one hour further away from the State capital and with a poverty rate higher by ten-percentage points.

Regarding school inputs, schools with test teachers have on average less students, but with a larger class size and are less likely to have a principal with graduate school training. As said before, Telesecundarias are small schools. Treatment schools have on average 87.4 students (9.85 less than those with discretionary teachers) distributed in around 4 classrooms (almost one less than the control schools).

The differences in terms of students' outcomes are less clear. There are not observed differences in the school scores in the Enlace examination -from here and after I will refer to grade 9 scores as school scores unless I specify otherwise. The share of students that are suspicious of cheating in this exam is larger in the treatment schools (10 percent versus 6 percent), a difference that is statistically significant at the five-percent level. In both type of schools, the number of students that take this end-of-the-year exam is on average 94 percent of those enrolled at the beginning of the school year.

As schools' location, inputs and outcomes are likely correlated, a regression analysis might 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 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 2 reports the results.

First, it is noteworthy that, holding constant school inputs and locality characteristics, no single

[18] As said before, I exclude here and after schools that received both test and discretionary teachers and schools that received a test teacher hired in 2008 and 2009.

measure of student performance in the last two years predicts assignment into treatment. In the same line, the p-value associated to the test of the joint significance of all the (past) student-performance variables included in the model is very high (.695) and hence we cannot reject the null hypothesis that they are jointly insignificantly different from zero at conventional levels of statistical significance. So, the data does not seems to support an assignment model in which joint committees allocate test teachers based on school (past) performance.

Second, class size is positively correlated with treatment status, while school size and the share of indigenous students in the school have a negative relationship with the probability of receiving a test teacher. However, the magnitude of the coefficients for class and school size (.00344 and -.00035) is relatively small -the difference in the mean values between treatment and control groups reported in Table 1 are 2.26 and -9.82 respectively).

Finally, it stands out the strong statistical relationship between treatment status and locality characteristics. Both the poverty rate and the adjusted distance to the state capital have coefficients that reflect a statistically and economically significant relationship with the probability of treatment assignment. In the same line, there is a strong negative relationship between the share of households with electricity service in the locality and treatment status. The coefficient for locality size has a positive sign and is statistically significant at the five percent level, although its magnitude is small. Summing up, test teachers are more likely to be allocated to schools in poorer and more isolated localities, with a lower penetration of public services like electricity.

Overall, the regressions analysis indicates that treatment status is strongly correlated with locality characteristics and, in a lesser degree, with school inputs. Also, there is no observed relationship between past school performance and the probability of assignment into treatment. These results are consistent with a model in which 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 is possible to control for both the effect of time-invariant locality characteristics and time-variant (observable) school inputs.

6 The Effect of Test-Hired Teachers

6.1 Identification and Estimation Methods

I am interested in estimating the average effect on student outcomes of assigning to a school a new teacher selected through a test-based examination (treatment) versus assigning a new teacher selected through a discretionary process with involvement from the teachers' union (control). With this purpose in mind, I estimate the following difference-in-differences model:

$$y_{st} = \beta_0 + \beta_1 T_{st} + \Gamma X_{st} + \tau_t + \alpha_s + v_{st} \tag{3}$$

Where y_{st} is an outcome of school s at time t, T is a indicator that equals 1 if school s received a test teacher (cohort 2010) at time t and 0 otherwise, β is the parameter of interest, X_{st} is a vector of time-variant school inputs and state effects and Γ is the associated vector of parameters, τ_t is a vector of year effects, α_s is a school time-invariant (at least for the period of interest) component and v_{st} is a disturbance term.

I estimate the model using panel data of Telesecundaria schools that received either test or discretionary teachers hired in the 2010 school year. I focus in Telesecundarias because their small size should increase the likelihood that I find a statistically significant teacher effect at the school level.

The estimation of a teacher effect at the school level has the drawback that teachers teach only one classroom in a Telesecundaria, while schools in the sample have on average almost 5 classrooms (4.8). Furthermore, I approximate school outcomes with grade 9th outcomes. I provide later an estimation of an scaled-up treatment effect.

Assuming conditional independence of T_{st}, the difference-in-differences parameter β captures the total (average) policy effect at the school level of allocating a teacher hired using a test-based examination versus on hired using a discretionary process with involvement from the teachers' union.

The total policy effect of using a recruiting method over other might comprise both: 1) the relative capacity of each method to identify and select teacher quality; and the propensity of (potential) candidates to apply through each of these methods. In other words, different sets of applicants might self-select into different recruitment methods. To what degree this happens or not, it is a question about the mechanisms through which the policy under study can relate to teacher quality. I focus for the moment on the identification of the total effect of the policy under study.

Though β is a relevant parameter from the policy point of view, it does not have the ceteris paribus interpretation of a parameter in the educational production as discussed by (*Todd and Wolpin*, 2003). Notably, there is no control for parental inputs that might react to changes in teacher quality induced by the policy. In principle, parents might increase or decrease the inputs they provide to students if they observe a change in teacher quality and teacher quality is a complement or a substitute of parental inputs.

As I discussed in section 2, the parameter β is informative about whether the discretionary process of teacher hiring with the involvement of the teachers' union follows a rent-seeking ($\beta > 0$) or an efficiency-enhancing ($\beta < 0$) model of participation.

The causal interpretation of β requires that the control schools give an accurate counter factual of the outcomes that the treatment schools would have had in the absence of treatment. Although it is impossible to test directly this assumption, I can test whether the secular trends in the treatment and control schools were the same in the pre-treatment period. Following *Galiani et al.* (2005), I estimate a modified version of equation 3 in which I use a fixed-effects model to regress the outcomes under study -in separate regressions- on a vector of year dummies interacted with (eventual) treatment status and a set of time-variant school inputs, time-variant state effects and year effects. I only use observations from the five years in the pre-treatment period. Table 3 reports the full-estimated model and results.

In the same line that results in Table 2, I do not find any statistically significant 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,[19] the share of students suspected of cheating in the school, and the school scores for Mathematics and Spanish in the Enlace exam. The coefficients for the interactions between (eventual) treatment status and year dummies in the regressions for final enrollment and share of flagged exams (columns 1 and 2) have all very small magnitudes and are not statistically significant either separately or jointly. In the regressions for the Mathematics and Spanish results (columns 3 and 4), the corresponding coefficients are sightlier larger, but only in one case the coefficient is statistically significant at the 10 percent-level. In all cases, including the former two, I cannot reject the null hypothesis that the pre-intervention year dummies are the same for both control and (eventual) treatment schools at conventional levels of statistical significance. I interpret these results as strong evidence in favor of the parallel trend assumption necessary for the identification of a causal average treatment effect in a difference-in-differences model.

6.2 Main Results: Binary Treatment

I present the main results of my difference-in-differences estimation in Table 4. The model in the four regressions controls for class size, school size, the share of indigenous students in the school, principal's attendance of graduate school and a vector of interactions between year and state dummies to capture state-specific time trends. Standard errors are clustered at the school level.

Regarding the results, I do not observe an effect of receiving (at least) a test teacher on the enrollment rate at the end of the academic year (see column 1). The coefficient of interest has a very small magnitude (.00166) and is not significant at conventional levels of statistical significance.

Notably, receiving a new teacher hired by the test examination reduces in 3.38 percentage points the share of students in the school who are flagged for exam cheating, which amounts to a 50 percent reduction of the baseline share. This result is statistically significant at the one-percent level.

As explained in section 4.1.1, the Federal Ministry 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. The Enlace regulations prohibit that teachers supervise the classroom they

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

teach during the exam application.

I interpret the negative effect of treatment on cheating as an indication of teacher quality. There are several reasons for why higher quality teachers can have a negative effect on exam copying. For example, better teachers could teach values -like academic honesty- that reduces students' propensity of cheating or they might exert more effort in the supervision of exams and in this way discourage exam copying. Also, higher quality teachers might be less likely to directly give answers to the students during the exam application.

I do not find that treatment affects students' test scores, measured through both the Mathematics and Spanish sections of the Enlace exam (columns 3 and 4). The magnitude of the estimated treatment effect on the Mathematics score is practically zero (-.0026 standard deviations) and statistically insignificant at the 10-percent level. While the magnitude of the corresponding coefficient is larger in the Spanish score case (-.0421 standard deviations), the associated standard error is relatively large (.0987) and the effect is not different from zero at conventional levels of statistical significance.

For interpretation, it is important to have in mind that I estimate results in columns 3 and 4 using the raw test scores. This neglects that the main purpose of cheating is to obtain higher scores. As the treatment reduces cheating, it is possible that there is actually a positive treatment effect on student achievement that I do not identify using the observed test scores.

6.3 Treatment Intensity

The model in equation 3 estimates an average effect on the treated for a binary treatment at the school level. There are reasons though for which this estimate might not be fully informative about the magnitude of the treatment effect under study. First, some schools actually received more than one new teacher in 2010. e.g. treated schools received on average 1.1 test teachers. Second, and more importantly, teachers teach only one classroom per school and I approximate school outcomes with grade 9th outcomes.

One could be interested in the average treatment effect on the treated at the classroom level, leaving aside within school externalities. I have already explained that I cannot directly estimate this treatment effect. However, I can approximate an estimate informative about the effect of teachers on the students they teach by modeling the intensity of the treatment at the school level.

I estimate a model like the one in equation 3 in which I substitute the indicator T_{st} for a variable S_{st} that measures the share of all teachers in the schools that were hired through the test examination in the school year 2010. Test teachers represent on average 32% of teachers in the treatment schools and 3.9% of teachers in all schools in the sample. Table 5 reports results.

As before, I do not find a treatment effect on the enrollment rate at the end of the academic year (column 1). Though the coefficient of interest has a negative sign, its magnitude is very small (.0082) and it is not different from zero at conventional levels of statistical significance.

I observe again that the allocation of test teachers has a strong and statistically significant effect on the probability of cheating (column 2). Going from not having test teachers in a school to having only test teachers is associated with a reduction of 6.99 percentage points in the share of students with flagged exams, an effect that is statistically significant at the one-percent level. This reduction is equivalent to the baseline level of detected cheating in all schools in the sample (6.74%).

As it is possible to see in columns 3 and 4, the magnitude of the treatment coefficients on the student achievement regressions is positive and considerably high (.28 standard deviation for Mathematics and .34 standard deviation for Spanish). However, the point estimates are very imprecise and are not significantly different from zero at conventional levels of statistical significance.

6.4 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 6.1 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 pre-treatment characteristics with unbalanced distributions between the treated and the control groups. For example, if the impact of test teachers is larger in isolated localities.

Abadie (2005) proposes a semi parametric difference-in-differences estimator -based on *Heckman et al.* (1997)- 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.

Following *Abadie* (2005), I estimate the propensity score from a logit model of the probability that a school receive a test teacher in the school year 2010 as a function of the pre-treatment characteristics listed in Table 2, although here I extend the use of lagged values to a period of five years before treatment. 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 information about outcomes one year before and one year after treatment. Results are reported in Table 6.

The picture emerging from Table 6 is consistent with those from Tables 4 and 5. First, the allocation of a test teacher does not have an effect on the share of students in the school that take the Enlace exam at the end of the academic year (column 1). Second, I observe again a large and highly statistically significant effect on the share of students flagged for copying in the final exam (column 3). The allocation of (at least) a test teacher is associated with a reduction of 8.81 percentage points -statistically significant at the one-percent level- in the share of students suspected of cheating. Finally, there is no observed effect on student test scores at conventional levels of statistical significance, although both coefficients of interest in columns 3 and 4 have a negative sign and a relatively large magnitude (-.16 and -.204 standard deviations, respectively).

6.5 Student Achievement and Test Cheating

I have presented results that show a strong negative relationship between treatment and the prevalence of exam copying and no significant relationship between treatment and student test scores. However, there are reasons to believe that my estimates of the treatment effect on student achievement have a downward bias. The treatment reduces cheating and the main purpose of cheating is to obtain higher scores. If cheating increases test scores, but not achievement, it is possible that there is actually a positive treatment effect on student achievement that I do not identify using the observed test scores.

There is indeed a strong correlation between exam copying and test scores in the data. In Table 7, I report the results of a model estimated in panel data with school fixed effects, in which I regress school test scores on the share of students in the school with flagged exams plus a vector of school inputs, year dummies and interactions of year dummies and state dummies. Both results in column 1 (Math Score) and column 2 (Spanish score) show a strong correlation between exam copying and test scores. Going from zero flagged exams in a school to only flagged exams is associated with an increase of 1.423 standard deviations in the Math score (1.31 for Spanish) statistically significant at the one-percent level.

As I would like to observe test scores in the absence of exam cheating, it might be useful to learn more about the school characteristics associated to cheating. With this purpose, I estimate a modified version of the linear probability model in Table 3, though here I use the share of flagged exams in the school in 2010 as the dependent variable and I include on the right-hand side of the equation a vector of dummies in which I categorize schools according to the mean of the share of flagged exams in the school during the pre-treatment period (2005-2009). The reference group is formed by the schools with no exam cheating detected during this period (49% of the sample).

It stands out that there is no statistically significant relationship between exam copying and any of the variables included for past student performance, school inputs and locality characteristics. On the contrary, there is a strong correlation between a school's history of exam copying and current exam copying. The coefficient for each of the dummy variables for schools with past exam copying are both economically and statically significant with respect to the schools with no past exam copying. This suggests that there is some structural factors -unobserved to me- that make exam cheating likely. Hence, estimating the treatment effect in a sample of schools with no recent history of exam cheating it might be informative about the effect of test teachers on student test scores in environments where cheating is rare.

I estimate first the same model than in Table 3 to investigate if in this restricted sample of schools with no previous history of exam cheating, treatment and control groups follow similar paths in the outcomes under study during the five years previous to treatment. The results from this placebo test are reassuring -see Table 9. The coefficients for the interactions between (eventual) treatment status and year dummies in the regressions for final enrollment, Math scores and Spanish scores have all very small magnitudes and are not statistically significant either separately or jointly at conventional levels of statistical confidence. Again, this evidence is consistent with a model in which teachers are allocated to schools based on locality characteristics and not according to trends in past performance.

I proceed then to estimate my difference-in-differences model for student outcomes -equation 3- in this group of schools with low propensity to exam copying. Results for the binary treatment specification are reported in Table 10. As before, I do not observe any treatment effect on the enrollment rate at the end of the academic year (column 1). More distinctly, I do not find either a statistically significant effect of allocating a test teacher on exam copying (column 2). The coefficient of interest still has a negative sign, but the magnitude is relatively small (-.01 percentage points) and is not significant at conventional levels of statistical confidence. On the contrary, treatment status is positively associated to both Math and Spanish test scores (columns 3 and 4). The sign for the coefficient of interest in both cases is positive and the magnitude is sizable (.13 and .142 standard deviations, respectively). The estimation is very imprecise though and the point estimates are not significantly different from zero at the ten-percent level.

Results for the specification with treatment intensity are reported in Table 11. As in the previous case, I do not find a statistically significant relationship between treatment status and the final enrollment rate and the share of flagged exams in the school (columns 1 and 2). I do find though an effect economically and statistically significant on test scores (columns 3 and 4). Moving from no test teachers in a school to only test teachers is associated with an increase in .663 standard deviations in the school's Mathematics test score (.788 for the Spanish score), a result that is statistically significant at the five-percent level (one-percent level for the Spanish score case).

Hence, in an environment in which exam copying is rare and in which test scores likely follow student achievement closer, the allocation of a test teacher has a positive and sizable effect on student achievement.

7 Conclusions

Summing up, I find that education officials hire on average teachers of less quality when they follow a discretionary process with participation from the teachers' union than when they rely on the results from a standardized test. Then, joint committees of state officials and union representatives allocate these teachers to schools in more "desirable" localities. Taken together, these results suggest the existence of an agency problem with potential rent-extraction.

In the same line, Duflo, Dupas and Kremer (2012) observe that weaker institutional settings in Kenya increase the probability that hiring committees hire a teachers' relative.

State officials are able to reduce the extent of the agency problem by using a hard-to-manipulate rule to hire new teachers -a ranking based on a teacher test- instead of relying on a discretionary process with strong participation from the teachers' union. Test teachers reduce exam cheating in the schools where they are allocated and -in environments where cheating is rare- increase (observed) student achievement.

These results are in contrast with the previous literature -mainly U.S. based- on the relationship between teacher scores and achievement tests -summarized in *Hanushek and Rivkin* (2006). The evidence emerging from this literature indicates that though teacher scores might be more informative about teacher quality than other teacher characteristics (like experience and education), they are still a modest predictor of teacher quality. Again, the interaction between the institutional context and the alternative hiring methods is likley key to determine the attractivenesss of teacher tests as a policy to identify teacher quality.

The findings described in this study are particularly relevant for environments with weak institutions, where the lack of accountability (on actual teacher performance, for example) may exacerbate the incentives for rent-seeking behavior arising because of imperfect information about worker productivity.

References

Abadie, A., Semiparametric difference-in-differences estimators, *Review of Economic Studies*, *72*(1), 1–19, 2005.

Duflo, E., P. Dupas, and M. Kremer, School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from kenyan primary schools, *NBER Working Papers 17939*, National Bureau of Economic Research, Inc, 2012.

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

Galiani, S., P. Gertler, and E. Schargrodsky, Water for life: The impact of the privatization of water services on child mortality, *Journal of Political Economy*, *113*(1), 83–120, 2005.

Hanushek, E. A., and S. G. Rivkin, *Teacher Quality*, *Handbook of the Economics of Education*, vol. 2, chap. 18, pp. 1051–1078, Elsevier, 2006.

Heckman, J. J., H. Ichimura, and P. E. Todd, Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme, *Review of Economic Studies*, *64*(4), 605–54, 1997.

Heckman, J. J., H. Ichimura, and P. Todd, Matching as an econometric evaluation estimator, *Review of Economic Studies*, *65*(2), 261–94, 1998.

Holland, P., Assesing unusual agreement between the incorrect answers of two examinees using the k-index, *Program Statistics Research*, 1996.

Hoxby, C. M., How teachers' unions affect education production, *The Quarterly Journal of Economics*, *111*(3), 671–718, 1996.

Jacob, B. A., and L. Lefgren, Can principals identify effective teachers? evidence on subjective performance evaluation in education, *Journal of Labor Economics*, *26*, 101–136, 2008.

Meyer, B. D., Natural and quasi-experiments in economics, *Journal of Business & Economic Statistics*, *13*(2), 151–61, 1995.

Rockoff, J. E., B. A. Jacob, T. J. Kane, and D. O. Staiger, Can you recognize an effective teacher when you recruit one?, *Education Finance and Policy*, *6*(1), 43–74, 2011.

Rothstein, J., Teacher quality in educational production: Tracking, decay, and student achievement, *The Quarterly Journal of Economics*, *125*(1), 175–214, 2010.

Santibanez, L., Reforma educativa: El papel del snte, *Revista Mexicana de Investigacion Educativa*, *13*(037), 419–443, 2008.

Todd, P. E., and K. I. Wolpin, On the specification and estimation of the production function for cognitive achievement, *Economic Journal*, *113*(485), F3–F33, 2003.

Figures

Table 1: Descriptive Statistics Across Schools with New Teachers (School Year 2009)

	Test Teachers	Discretionary Teachers	diff.
Math Score Grade 9 (std)	-0.00	-0.05	0.04
	(0.10)	(0.02)	(0.07)
Spanish Score Grade 9 (std)	-0.01	-0.01	0.00
	(0.10)	(0.03)	(0.08)
Flagged Exams Grade 9	0.10	0.06	0.04**
	(0.02)	(0.01)	(0.02)
Final Enrollment Grade 9th	0.94	0.94	-0.00
	(0.01)	(0.00)	(0.01)
Students	87.42	97.27	-9.85
	(5.26)	(2.71)	(7.43)
Class size	19.83	17.58	2.26***
	(0.80)	(0.23)	(0.67)
Groups	4.06	4.94	-0.88***
	(0.14)	(0.09)	(0.24)
Share indigenous students	0.19	0.14	0.05*
	(0.03)	(0.01)	(0.03)
Principal has grad school	0.20	0.27	-0.07*
	(0.03)	(0.01)	(0.04)
Locality Population	9985.71	45931.64	-35945.93**
	(5278.64)	(6155.70)	(16398.21)
Hours to state capital	2.92	2.08	0.84***
	(0.16)	(0.05)	(0.14)
Locality Poverty Rate	0.76	0.66	0.10***
	(0.01)	(0.01)	(0.01)
Share hhs electricity	0.89	0.94	-0.05***
	(0.02)	(0.00)	(0.01)
Share hhs sewage	0.64	0.70	-0.05**
	(0.03)	(0.01)	(0.03)
Observations	155	1090	1245

* p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

Significance levels are reported for t-tests of the equality of the means across treatment status

Table 2: Probability of Receiving a Test Teacher (OLS)

VARIABLES	(1) School Year 2010
Math Score Grade 9th Lag 1	-0.0137
	(0.0198)
Spanish Score Grade 9th Lag 1	6.37e-05
	(0.0187)
Flagged exams grade 9 Lag 1	0.0145
	(0.0635)
Final Enrollment Grade 9th Lag 1	-0.0396
	(0.115)
Math Score Grade 9th Lag 2	0.0330
	(0.0233)
Spanish Score Grade 9th Lag 2	-0.0136
	(0.0205)
Flagged exams grade 9 Lag 2	-0.0250
	(0.0793)
Final Enrollment Grade 9th Lag 2	0.107
	(0.0731)
Class size Lag 1	0.00344*
	(0.00199)
Students Lag 1	-0.000354***
	(0.000123)
Share indigenous students Lag 1	-0.0693**
	(0.0344)
Principal has grad school	-0.0146
	(0.0189)
Locality Poverty Rate	0.159**
	(0.0768)
Hours to state capital	0.0249***
	(0.00921)
Locality Population	4.92e-08*
	(2.80e-08)
Share hhs electricity	-0.448***
	(0.117)
Share hhs sewage	0.00220
	(0.0400)
Constant	0.291*
	(0.161)
Observations	1,124
R-squared	0.192
State Fixed Effects	Yes
F statistic Ho Var 1-8=0	0.685
Prob>F	0.705

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Difference-in-Differences: Pre-Treatment Trends in Outcomes

VARIABLES	(1) Final Enrollment	(2) Share Flagged Exams	(3) Math Score (std)	(4) Spanish Score (std)
2006 X Test Teacher	0.00160	0.0164	0.121	0.0686
	(0.0143)	(0.0134)	(0.108)	(0.101)
2007 X Test Teacher	0.00213	-0.0193	0.0653	0.0886
	(0.0142)	(0.0209)	(0.110)	(0.0989)
2008 X Test Teacher	0.0277*	0.0104	0.240*	0.189
	(0.0163)	(0.0200)	(0.124)	(0.119)
2009 X Test Teacher	0.00967	0.00254	0.147	0.178
	(0.0135)	(0.0194)	(0.121)	(0.121)
Class size	-0.00105	-0.000414	-0.00673	-0.00300
	(0.000770)	(0.00113)	(0.00588)	(0.00565)
Students	-0.000279*	0.000230	-0.000499	-0.000652
	(0.000168)	(0.000189)	(0.00116)	(0.000955)
Share indigenous students	-0.0111	-0.00111	0.0222	0.100
	(0.0114)	(0.0154)	(0.0827)	(0.0844)
Principal has grad school	0.0117**	0.00122	0.00405	-0.0450
	(0.00515)	(0.00942)	(0.0477)	(0.0461)
Observations	5,984	5,986	5,985	5,985
R-squared	0.075	0.050	0.039	0.051
Number of id	1,231	1,231	1,231	1,231
School Fixed Effects	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
State X Year Dummies	Yes	Yes	Yes	Yes
F statistic Ho Var 1-4=0	1.093	0.920	1.060	0.818
Prob>F	0.358	0.451	0.375	0.514

Grade 9th Outcomes. Robust standard errors in parentheses clustered at the school level
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Difference-in-Differences: Results (Binary Treatment)

VARIABLES	(1) Final Enrollment	(2) Flagged Exams (%)	(3) Math Score (std)	(4) Spanish Score (std)
2010 Test Teacher	0.00338	-0.0327**	-0.0238	-0.0520
	(0.00875)	(0.0132)	(0.0963)	(0.0971)
Class size	-0.000580	0.000624	-0.00858	-0.00657
	(0.000613)	(0.000902)	(0.00529)	(0.00525)
School size	-0.000204	4.55e-05	-0.000486	-0.000476
	(0.000127)	(0.000141)	(0.000970)	(0.000886)
Share indigenous students	-0.00564	0.00489	-0.0134	0.0481
	(0.00938)	(0.0133)	(0.0819)	(0.0831)
Principal has grad school	0.00935**	0.00162	0.00884	-0.0486
	(0.00424)	(0.00751)	(0.0437)	(0.0427)
Constant	0.944***	0.000813	0.193**	0.165*
	(0.0114)	(0.0143)	(0.0926)	(0.0917)
Observations	7,210	7,212	7,211	7,211
R-squared	0.074	0.044	0.051	0.074
Number of id	1,232	1,232	1,232	1,232
School Fixed Effects	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
State*Year Dummies	Yes	Yes	Yes	Yes

Grade 9th Outcomes. Robust standard errors in parentheses clustered at the school level
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Difference-in-Differences: Results (Intensity Treatment)

VARIABLES	(1) Final Enrollment	(2) Flagged Exams (%)	(3) Math Score (std)	(4) Spanish Score (std)
Share Test Teachers	-0.00580	-0.0660**	0.217	0.255
	(0.0332)	(0.0267)	(0.260)	(0.254)
Class size	-0.000590	0.000689	-0.00844	-0.00635
	(0.000612)	(0.000905)	(0.00530)	(0.00525)
School size	-0.000201	2.92e-05	-0.000522	-0.000535
	(0.000127)	(0.000141)	(0.000970)	(0.000888)
Share indigenous students	-0.00559	0.00488	-0.0144	0.0467
	(0.00938)	(0.0133)	(0.0817)	(0.0830)
Principal has grad school	0.00930**	0.00172	0.00963	-0.0474
	(0.00424)	(0.00752)	(0.0437)	(0.0428)
Constant	0.944***	0.00115	0.193**	0.167*
	(0.0114)	(0.0143)	(0.0925)	(0.0917)
Observations	7,210	7,212	7,211	7,211
R-squared	0.074	0.044	0.051	0.074
Number of id	1,232	1,232	1,232	1,232
School Fixed Effects	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
State*Year Dummies	Yes	Yes	Yes	Yes

Grade 9th Outcomes. Robust standard errors in parentheses clustered at the school level
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Difference-in-Differences with Propensity Score Matching: Results (Binary Treatment)

VARIABLES	(1) Final Enrollment (%)	(2) Flagged Exams (%)	(3) Math Score (std)	(4) Spanish Score (std)
Test	-0.00310	-0.0881***	-0.160	-0.204
	(0.0162)	(0.0320)	(0.232)	(0.238)
Observations	863	863	863	863

Grade 9th Outcomes. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Correlation between Test Scores and Exam Cheating

VARIABLES	(1) Math Score (std)	(2) Spanish Score (std)
Flagged Exams (%)	1.423***	1.310***
	(0.105)	(0.0990)
Class size	-0.00962*	-0.00745
	(0.00510)	(0.00510)
Students	-0.000527	-0.000534
	(0.000928)	(0.000863)
Share indigenous students	-0.0190	0.0427
	(0.0793)	(0.0785)
Principal has grad school	0.00655	-0.0504
	(0.0414)	(0.0410)
Constant	0.194**	0.167*
	(0.0893)	(0.0892)
Observations	7,211	7,211
R-squared	0.112	0.125
Number of id	1,232	1,232
School Fixed Effects	Yes	Yes
Year Dummies	Yes	Yes
State*Year Dummies	Yes	Yes

Grade 9th Outcomes. Robust standard errors in parentheses clustered at the school level
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Probability of Exam Cheating (OLS)

VARIABLES	(1) 2010 Flagged Exams (%)
2005-09 Mean Flagged Exams 0.01%-10%	0.0265***
	(0.00837)
2005-09 Mean Flagged Exams 10%-20%	0.0449*
	(0.0234)
2005-09 Mean Flagged Exams 20%-100%	0.0899***
	(0.0257)
Math 9 Lag 1	0.00695
	(0.00948)
Spanish 9 Lag 1	0.00733
	(0.0100)
Final Enrollment Grade 9 Lag 1	0.00581
	(0.0491)
Math 9 Lag 2	-0.00270
	(0.00800)
Spanish 9 Lag 2	-0.00280
	(0.00773)
Final Enrollment Grade 9 Lag 2	-0.0283
	(0.0403)
Class size Lag 1	0.000195
	(0.000771)
Students Lag 1	-6.60e-05
	(4.75e-05)
Share indigenous students Lag 1	0.00708
	(0.0120)
Principal has grad school Lag 1	-0.000230
	(0.00987)
Locality Poverty Rate	-0.0530
	(0.0340)
Hours to state capital	-0.000132
	(0.00304)
Locality Population	-5.76e-09
	(1.69e-08)
Share hhs electricity	0.0263
	(0.0259)
Share hhs sewage	-0.00778
	(0.0171)
Constant	0.0725
	(0.0548)
Observations	1,118
R-squared	0.084
State Fixed Effects	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Difference-in-Differences: Pre-Treatment Trends - Restricted Sample

VARIABLES	(1) Final Enrollment (%)	(2) Math Score (std)	(3) Spanish Score (std)
2006 X Test Teacher	-0.000746	0.0255	-0.00180
	(0.0222)	(0.136)	(0.140)
2007 X Test Teacher	0.000949	-0.0444	-0.0483
	(0.0220)	(0.137)	(0.146)
2008 X Test Teacher	0.0415	0.220	0.0258
	(0.0266)	(0.168)	(0.177)
2009 X Test Teacher	0.00982	-0.133	-0.214
	(0.0216)	(0.162)	(0.173)
Class size	-0.00182	-0.00595	0.00242
	(0.00113)	(0.00880)	(0.00924)
Students	-0.000420	-0.00280*	-0.00154
	(0.000307)	(0.00151)	(0.00168)
Share indigenous students	-0.0274	-0.0610	0.0172
	(0.0172)	(0.0844)	(0.101)
Principal has grad school	0.0186**	-0.0790	-0.0993
	(0.00849)	(0.0689)	(0.0669)
Observations	2,965	2,965	2,965
R-squared	0.085	0.061	0.050
Number of id	610	610	610
School Fixed Effects	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes
State X Year Dummies	Yes	Yes	Yes
F statistic Ho Var 1-4=0	0.975	1.714	0.906
Prob>F	0.420	0.145	0.460

Grade 9th Outcomes. Robust standard errors in parentheses clustered at the school level
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Difference-in-Differences: Results - Restricted Sample (Binary Treatment)

VARIABLES	(1) Final Enrollment (%)	(2) Share Flagged Exams (%)	(3) Math Score (std)	(4) Spanish Score (std)
2010 Test Teacher	0.0138	-0.0109	0.130	0.142
	(0.0137)	(0.0112)	(0.117)	(0.120)
Class size	-0.000634	0.000631	-0.00107	0.00482
	(0.000932)	(0.000542)	(0.00751)	(0.00839)
Students	-0.000423*	-0.000188*	-0.00433***	-0.00321**
	(0.000254)	(0.000107)	(0.00143)	(0.00159)
Share indigenous students	-0.0146	0.0114	-0.0682	-0.0139
	(0.0139)	(0.0106)	(0.0896)	(0.0962)
Principal has grad school	0.0138*	-0.00112	-0.0459	-0.0742
	(0.00713)	(0.00269)	(0.0614)	(0.0592)
Constant	0.956***	0.00153	0.246**	0.165
	(0.0149)	(0.00433)	(0.0975)	(0.111)
Observations	3,572	3,572	3,572	3,572
R-squared	0.083	0.074	0.077	0.096
Number of id	611	611	611	611
School Fixed Effects	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
State*Year Dummies	Yes	Yes	Yes	Yes

Grade 9th Outcomes. Robust standard errors in parentheses clustered at the school level
*** p<0.01, ** p<0.05, * p<0.1

Table 11: Difference-in-Differences: Results - Restricted Sample (Intensity Treatment)

VARIABLES	(1) Final Enrollment (%)	(2) Share Flagged Exams (%)	(3) Math Score (std)	(4) Spanish Score (std)
Share Test Teachers	0.0162	-0.0228	0.663**	0.788***
	(0.0488)	(0.0197)	(0.306)	(0.272)
Class size	-0.000667	0.000653	-0.00121	0.00470
	(0.000930)	(0.000557)	(0.00748)	(0.00834)
Students	-0.000411	-0.000197*	-0.00429***	-0.00318**
	(0.000252)	(0.000111)	(0.00142)	(0.00157)
Share indigenous students	-0.0146	0.0114	-0.0696	-0.0156
	(0.0138)	(0.0106)	(0.0894)	(0.0961)
Principal has grad school	0.0137*	-0.00111	-0.0442	-0.0721
	(0.00713)	(0.00268)	(0.0614)	(0.0591)
Constant	0.955***	0.00173	0.245**	0.164
	(0.0149)	(0.00427)	(0.0973)	(0.111)
Observations	3,572	3,572	3,572	3,572
R-squared	0.082	0.073	0.080	0.099
Number of id	611	611	611	611
School Fixed Effects	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
State*Year Dummies	Yes	Yes	Yes	Yes

Grade 9th Outcomes. Robust standard errors in parentheses clustered at the school level
*** p<0.01, ** p<0.05, * p<0.1