Productivity and Sectoral Allocation: The Labor Market of School Principals

Job Market Paper
For most recent version, please click here *

Pablo Muñoz Mounu Prem

Abstract

An important question is whether differences in management can explain variation in productivity and whether different compensation schemes and selection policies can impact the allocation of managerial effectiveness. To investigate this, we measure the value-added of school principals and study their labor market outcomes in Chile. Using large administrative data sets on student performance and school personnel, we document substantial variation across principals in their ability to improve students' learning. We show that effective principals increase the retention of young and high value-added teachers, and that principal effectiveness is recognized by the school community. A theoretical model of job offers and acceptances guides our empirical inquiry into the principals' labor market. Leveraging observational and quasi-experimental variation, we show that: (1) principals' effectiveness is more strongly associated with wages in private schools, and (2) despite relatively rigid wages, public schools can attract better principals by improving their personnel selection process.

*This version: October 30, 2019. Corresponding author: Pablo Muñoz, University of California at Berkeley, Department of Economics. Email: pablomh@berkeley.edu. We are particularly indebted to David Card, Fred Finan, and Jesse Rothstein for their advice and support at different stages of this project. We are also indebted to Ernesto Dal-Bó, Alonso Bucarey, Christopher Campos, Dante Contreras, Patrick Kline, Mathieu Pedemonte, Tatiana Reyes, Damian Vergara, and Christopher Walters, as well as participants at the UC Berkeley Labor Seminar and Political Economy Research Lunch for helpful suggestions. Our thanks to the Superintendencia de Educación and the Unidad de Estadísticas del Ministerio de Educación for their help with the data. The views expressed here are those of the authors and do not reflect the views of these institutions. We are grateful to the Spencer Foundation for financial support. José Niño and Luis Serrano provided excellent research assistance. Mounu Prem, Universidad del Rosario, Department of Economics. Email: francisco.munoz@urosario.edu.co

1 Introduction

While it is beyond dispute that the quality of the workforce is a key resource of both private and public enterprise (e.g., Lazear and Shaw, 2007; Finan et al., 2015), only recently have scholars begun to gather observational estimates of worker-specific productivity. Moreover, despite growing evidence that management practices matter (e.g., Bloom et al., 2013), little is known about the productivity of workers in managerial positions and its relation with labor market outcomes. In particular, there is scant evidence on whether managers' productivity can be recognized by their workers and labor market; and whether differences in personnel selection and wage schemes can help explain the allocation of managerial effectiveness in the public and private sectors.

In this article, we study the productivity of school principals and its relation to their labor market outcomes. As managers of the schools, principals are responsible for multiple tasks such as setting goals, reviewing students' and teachers' progress, keeping a safe school climate, and being a hub for interaction among teachers, students, and parents. We focus on Chile, where public and private schooling sectors have coexisted and competed in the same market for almost forty years, with public schools serving more unprivileged children and exhibiting worse outcomes than their private counterparts.[1] Like many countries, Chilean public and private schools have different compensation schemes, i.e., public schools' wages are heavily regulated and mostly based on seniority, while private schools have more flexibility to make personnel decisions and to set wages. Unlike many countries, in Chile it is possible to systematically link administrative employer-employee records of the universe of teachers and principals to the achievement of their students, making it possible to construct objective measures of performance.

We develop a novel extension of the canonical teacher value-added model (VAM) that can be used to measure principals' effectiveness. We find substantial variation across principals in their ability to improve students' learning. A one standard deviation increase in princi-

[1]Evidence on whether public or private schools are relatively more effective in improving students' learning is mixed (e.g., Hsieh and Urquiola, 2006; Contreras et al., 2018), probably due to difficulties in accounting for unobserved factors determining the selection of children attending each type of school.

pal effectiveness raises students' course grades by 0.17 standard deviations.[2] We use these effectiveness measures to investigate various dimensions of principals' management and labor market outcomes, with special attention to the differences between private and public schools. We obtain three main findings. First, effective principals increase the retention of young and high value-added teachers. Second, according to administrative records, effective principals are recognized by teachers and the school community, and their effectiveness is compensated in the labor market, especially by private schools. Third, despite a rigid wage scheme, public schools can attract better principals by improving their personnel selection process.

The first part of our study measures principal effectiveness. We estimate a value-added model using more than 8.5 million student-subject-year observations. This model relates academic achievement to student and school characteristics, and to more than 9,000 principal fixed effects. An important challenge in the estimation of principals' effects, that does not apply in teacher value-added modeling, is the need to disentangle the effect of the principal from other school-level factors. To address this issue, we follow the seminal work by Mundlak (1978), and the more recent work of Altonji and Mansfield (2018), and include several school-level controls in a correlated random effects fashion. This approach not only allows us to obtain an estimate of effectiveness for all school principals, but also overcomes problems of weak identification that could arise with a two-way school-principal fixed effects model (e.g., Jochmans and Weidner, 2019). A related challenge is distinguishing the principals' effectiveness from the average effectiveness of the school's teachers. Since teachers' effectiveness is itself a potential mechanism by which principals influence student achievement, we do not take teacher effects into account directly. Instead, we use the estimated variance of the error components to construct *shrunken* empirical Bayes posterior estimates of teacher effectiveness, as in Kane and Staiger (2008).

We use our measure of principal effectiveness to investigate three main questions. First, do better principals increase teacher retention? Although teacher shortages and high turnover rates have recently received considerable attention (e.g., Ronfeldt et al., 2013; Hanushek et al., 2016), evidence on the role of school managers has been lacking. To investigate this, we use a rich employer-employee panel including teacher and school characteristics from 2008 to 2016. We find that a one standard deviation increase in principal quality decreases

[2] A non-trivial effect size in the Chilean context, where a one standard deviation increase in GPA is associated to an 11 percent increase in wages, controlling for test scores, school characteristics, and students demographics.

the average exit rate of teachers by 3%, with stronger effects for young and high value-added teachers. A one standard deviation increase in principal effectiveness decreases teacher turnover by 5% among young teachers (between 22 and 37 years old) and by 7% among high value-added teachers (in the top tercile of the empirical value-added distribution). These estimates are robust to specifications with teacher fixed effects, and they imply that teachers are willing to pay around 3% of their wages for a one standard deviation better principal.[3]

The second question motivating our study is whether teachers, the school community, and the labor market can recognize effective principals. This is relevant to assess whether our measure of principals' effectiveness is sensible. Using teachers' surveys, we find that more effective principals can be recognized, i.e., teachers working for them are more likely to agree with statements such as the principal *does a good job*, *promotes a good work climate*, and *can be trusted*. Effective principals also receive fewer complaints from the school communities. A one standard deviation increase in principals' effectiveness is associated with a decrease of complaints about denied enrollment and bullying by an 18% and 17%, respectively. These results are robust to permutation tests and to the adjustments for multiple testing proposed by Romano and Wolf (2005). Regarding the principals' labor market, we show that principals' productivity is rewarded, especially by private schools. A one standard deviation increase in principals' effectiveness is associated with 4% higher wages in private schools, but only 1% in public schools.[4] This result is consistent with two stylized facts of our data: public schools rely less on pay-for-performance and wages in private schools exhibit more variance than wages in public schools.

The equilibrium relationship between principals' quality and wages motivates our third and final question: can public schools attract high-quality principals despite having more rigid wages? As we show through a two-sided selection model, if higher quality candidates demand higher compensation, then higher wages are a necessary condition for attracting those candidates. However, higher wages might not suffice nor be the only relevant variable as workers' choice also depends on their idiosyncratic preferences and the labor demand that they face, i.e., school principals can only choose among the job offers that are available to them. As highlighted by economists and sociologists in the past (e.g., Abowd and Farber, 1982; Logan, 1996), when choice is constrained by opportunity, personnel selection can ac-

[3] We can estimate this willingness to pay using the known negative correlation between wages and labor turnover (Manning, 2011). We find a significant effect of wages on teacher turnover, which implies an average turnover elasticity of -0.7, half of the one reported by Hendricks (2014).

[4] We also find that tenure is more relevant in public than private schools and that there exists a gender gap against female principals, which is accentuated in private schools.

centuate or counteract the sorting of workers based on the payment component. Moreover, poor selection policies may lead public schools to a loss of talented and motivated workers who would take a position despite a low payment because of strong idiosyncratic preferences for the public sector (e.g., pro-social behavior). Although the selection dimension of the labor market is of high practical relevance, empirical progress has met with at least two important hurdles. First, it is difficult to measure an individual's quality. Second, it is hard to find quasi-experimental variation in personnel selection policies.

In this study, we overcome these limitations by exploiting our measure of principals' effectiveness and the features of a reform that changed the selection process of public school principals. In 2011, Chile enacted the Law 20.501 of *Quality and Equity in Education*. This Law comprised a set of policies intended to level conditions between public and private schools, with a special focus devoted to the improvement of leadership in public schools. Before the reform, the appointment of a new public school principal was the exclusive responsibility of the municipalities; this process was unsupervised by the central government and was therefore more prone to patronage (e.g., Akhtari et al., 2018; Colonnelli et al., 2019). Under the new system, principals must be elected through public, competitive, and transparent competitions in a process that is lead by a third-party human resources agency and that is overseen by the Civil Service agency responsible for selecting the highest-level bureaucrats of the central government. Leveraging the timing of the new selection system's adoption, we implement a staggered difference-in-differences approach to assess the effects of this policy on the average quality of principals in public schools. Our estimates suggest that this new selection process increased principal effectiveness in public schools by 0.04 standard deviations, enough to close one-half of the public-private achievement gap in 6 years.

In Chile, students do not take nationwide standardized tests every year; thus we measure students' performance in Mathematics and Spanish using their course grades.[5] While we show that our results are robust to an alternative measure of principal effectiveness based on test scores, we argue that course grades are well suited to our purposes for the following reasons. First, Chile has a standardized national curriculum and consequently test scores and course grades are strongly positively correlated, supporting the view of using them interchangeably (e.g., Borghans et al., 2016). Second, as is common in value-added models, we examine performance longitudinally, which should mitigate concerns about grade inflation.

[5] As in Peteck and Pope (2019), we address the concern that teachers can directly impact grades by using the lead of course grades as the outcome variable, and we estimate the model on a sample of students who changed their teacher from one year to another.

Third, grades also capture non-cognitive traits, which are valued by the labor market (e.g., Bowles and Gintis, 1976; Bowles et al., 2001; Heckman et al., 2006). Indeed, course grades —and not test scores— determine grade retention and high school graduation. Fourth, and finally, estimates of principal effectiveness are more precisely estimated using course grades than test scores, as we observe the former every year for all students.

Related Literature

Related research on school principals has focused on the effect of principal attributes on students' performance (e.g., Eberts and Stone, 1988; Clark et al., 2009; Béteille et al., 2012) and on the measurement of principal effectiveness by itself (e.g., Coelli and Green, 2012; Dhuey and Smith, 2014; Grissom et al., 2015). A few studies have also looked at the relationship between principals and teachers (e.g., Harris and Sass, 2006; Jacob and Lefgren, 2008) and at the labor market of school principals (e.g., Cullen et al., 2016). In this context, our research offers a broader, unified study of school principals, and, in doing so, it contributes to a number of different branches of the economics literature.

First, our work relates to a large body of research estimating and discussing the validity of value-added models (e.g., Kane and Staiger, 2008; Rothstein, 2010; Kane et al., 2013; Chetty et al., 2014; Bacher-Hicks et al., 2014; Bacher-Hicks et al., 2014; Rothstein, 2015; Chetty et al., 2016; Angrist et al., 2017). In contrast to most papers that focus on the effects of teachers and schools on student outcomes, we contribute new evidence to an incipient literature that looks at the effectiveness of school principals (Branch et al., 2012; Coelli and Green, 2012; Dhuey and Smith, 2014; Grissom et al., 2015). Specifically, we extend the canonical value-added model with teacher random effects by including school correlated random effects and principal fixed effects; we also address concerns about bias in VAM by evaluating the validity of our estimated measure of principal's effectiveness.

Second, our paper also relates to recent research on the importance of management and management practices in the private sector (e.g., Bloom and Van Reenen, 2007; Bloom et al., 2013; Bender et al., 2018) and public sector institutions (e.g., McCormack et al., 2014; Bloom et al., 2015; Lavy and Boiko, 2017; Rasul and Rogger, 2018; Fenizia, 2019). By attempting to understand what effective principals do (and do not), our work contributes to research on teacher turnover (e.g., Hanushek et al., 2004; Ransom and Sims, 2010; Ronfeldt et al., 2013; Hendricks, 2014; Hanushek et al., 2016) and student tracking (e.g., Duflo et al., 2011; Card and Giuliano, 2016).

Finally, our analysis of the labor market of school principals relates to classical studies on the relationship between workers' wage and productivity (e.g., Marshall, 1890; Hicks, 1963; Robinson, 1969; Lazear and Rosen, 1981), and more specifically to research on personnel economics and the public personnel of the State (see Lazear and Shaw, 2007; and Finan et al., 2015 for reviews). On one hand, our study relates to literature on school personnel (e.g., Rothstein, 2015; Biasi, 2018; Loyalka et al., 2019), as we study the relationship between principal effectiveness and principal compensation. On the other hand, our analysis using quasi-experimental variation is close to that of Ferraz and Finan (2011) and Dal Bó et al. (2013), who in different settings show that competition and compensation can help to improve state capabilities.

The rest of this paper is organized as follows. The next section discusses institutional features of the Chilean educational system and presents the data and descriptive statistics. Section 3 explains and discusses the main empirical strategy used to estimate principal effectiveness. Section 4 explores the relationship between principal effectiveness, school management, and perceptions of the school community. Section 5 studies whether principal effectiveness is recognized by the labor market, and section 6 presents a theoretical model to think about principals' choice given available job offers. Section 7 shows the effect that the new selection system had on public schools, and section 8 concludes.

2 Institutional Setting and Data

This section discusses the main educational reforms implemented in Chile since the eighties and the current functioning of the educational system. It also describes the data used in our analysis and presents some descriptive statistics.

Historical Background

In 1981, under a dictatorship, Chile implemented an educational reform that privatized and decentralized primary and secondary education. Publicly funded school vouchers were created with flat voucher funds following any children either to public schools or to the private schools that agreed to accept the voucher as payment of tuition.[6] These vouchers provided full coverage of tuition fees in public schools, but not necessarily in private subsidized schools, which were allowed to charge fees on top of the part covered by the voucher. Indeed, while public schools could not profit, private schools that agreed to accept the vouchers could do it (Elacqua, 2009). The reform was predicated upon the idea that, since parents were free to choose between schools, market forces should lead to an increase in the quality of education through school competition. The *laissez faire* architecture of this system is still in place today, making the Chilean case unique for having long-term experience with nationwide school vouchers where both governmental and private schooling sectors coexist and compete.[7] As of 2018, enrollment at private, subsidized private, and public schools represented a 7, 53, and 40 percent respectively.

Alongside the privatization of the educational system, the 1981 reform also decentralized it by transferring the control of public schools from the central government to municipal authorities. Administrative departments of municipal education (DAEM by its Spanish acronym) and municipal education corporations were created to administrate the public schools. The efficacy and probity of these departments/corporations was strongly related to that of the local governments (Guerra and Arcos, 2012). As a consequence of this change, many school teachers from public schools lost their jobs and had to either reapply for them, now in the municipal sector, or to find jobs in the private sector. Moreover, in order to *free* the labor market of teachers, union contracts were revoked, giving public schools greater flexibility in hiring and firing teachers.

After the restoration of democracy in 1990, the teacher's association was reinstated and teacher pay was increased. Public school teachers once again belong to a national teachers union, while those in private schools are generally members of school-level teachers' associations. During the period we study, teachers' wages in the public sector were determined by a rigid formula that was negotiated between the government and the national teachers union. Wages are subject to seniority increments, and other adjustments such as allowances for leadership responsibilities, professional training, and for working in difficult conditions. Teachers in private schools are also eligible for some of these allowances, but they are mainly subject

[6] A reform in 2008 (Law 20.248) established a new targeted voucher for each eligible low-income student. This effectively eliminated out-of-pocket tuition paid by eligible students at participating schools because it forced schools to not add any additional top-up fees. This represented the first major change to the voucher policy program. Eligibility to the program was reserved for approximately the poorest 40% of the population, as determined by the central government. For an evaluation of this policy, see Neilson (2019).

[7] The map presented in figure C.1, in the appendix, exemplifies the coexistence of public and private schools in a low-income municipality at the north of Santiago.

to the Private Labor Code, implying that their wages are often individually negotiated with the schools.

In 2011, and after massive students' protests, the country enacted a reform aimed at improving quality and equity in education (Law 20.501). This reform comprised a set of policies intended to level conditions between public and private schools, with a special focus devoted to improve the leadership in public schools. This law created a new process of selection and increased the attributions of school principals. After the reform, all new principals of public schools i) must be elected in a public and competitive contest run by an external human resources firm and overseen an by the same agency responsible for selecting the highest-level bureaucrats of the central government; ii) are allowed to form their own management teams without having to call a contest for those positions, i.e., they can choose the Deputy Director, the Inspector General and the Chief Technician of the school; iii) can fire up to a 5% of teachers with a bad or regular teacher evaluation; and finally iv), consistent with their higher responsibilities and the leadership they must exercise in their position, principals at public schools get an special bonus in accordance to a rule that depends on the total number of students enrolled and the concentration of poor students in the establishment where they work.[8]

The Current Educational System

Compulsory education in Chile consists of 12 years divided into two cycles: primary school from 1st to 8th grade and secondary school from 9th to 12th grade. Most of the students enter the educational system when they are six years old and finish when they are eighteen. On average, each of the 9,000 schools in the country enrolls about 300 students, excluding preschools, adults' schools, and special education schools. According to the Chilean Ministry of Education, 46% of the schools are in areas considered rural, and 49% of the students are eligible for a special subsidy targeted to the poor.

Like many countries, Chile has a nationwide standardized curricula.[9] The Chilean curriculum

[8] The reform included other measures as well, such as a new retirement plan for teachers, bonuses for teachers with good evaluations, adjustments to the severance payments, introduction of public contests for the position of educational superintendent, more rights for teachers and teacher assistants, and more funding for both public and subsidized private schools. For details see "Ley 20.501 Calidad Y Equidad de la Educación", available at https://www.leychile.cl/Navegar?idLey=20501 (last accessed on July 2nd, 2019).

[9] Countries that have a national curriculum include: France, Hungary, Ireland, Italy, Japan, Korea, the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, and the United Kingdom. While most

is determined by the Ministry of Education for each grade and subject, and it affects the school curricular offerings and the instructional resources directly. It also works as a system of accountability (Valverde, 2004). Indeed, as exemplified by figure C.2, the government not only provides teachers with the curriculum guides and official textbooks, but also with lesson plans and exams. Finally, the curriculum guidelines also establish minimum content goals and fundamental objectives for education, which ultimately determine course grades and grade retention.

Teachers in a particular subject determine the course grade in that subject. Students are evaluated continuously throughout the year, and, in general, each subject's annual mark is based on more than four evaluations. Marks are graded on a scale from 1 to 7 in intervals of 0.1, with a minimum passing mark of 4. The Ministry of Education also administers a national standardized test (SIMCE by its acronym in Spanish). This test comprehend a subset of subjects (Math, Spanish, History, Science and English), and it is taken annually but only by students in the 4th, 8th, and 10th grades.

Regulations for the operation and organisation of schools differ between the public and the private-subsidised sectors, thus giving school providers and schools in both sectors different levels of autonomy to hire, pay, and administrate the schools (Paulo et al., 2017).[10] However, school principals in both systems need to have relevant training and teaching experience to be eligible for the position. According to the Chilean Law, all school principals should at least: i) develop, monitor, and evaluate the goals and objectives of the school; ii) study strategies for the implementation of the national curricula; iii) organise and guide the technical-pedagogical work and professional development of teachers; and iv) ensure that parents or guardians receive regular information on the operation of the school and the progress of their children.

Data and Descriptive Statistics

We use administrative data from the Ministry of Education, the Superintendency of Education, and the Chilean Civil Service.

To estimate principals' effectiveness, we use a panel at the student-year-subject level. This

States in the U.S. follow common guidelines for a core curriculum, there is no national curriculum as such.

[10]Unlike subsidized private schools, public schools are subject to a more restrictive labor regulation that specifies procedures for recruitment, dismissal, and remuneration. These schools also have to comply with special regulations regarding the purchase of goods and services.

panel spans from 2011 to 2016 and has information on the academic performance of all students in public and subsidized private schools, by subject and classroom, from the first through the twelfth grade. Specifically, the students' records contain their gender, age, subject-specific course grades, attendance rate, and promotion status. For cohorts of students that take standardized exams, it is also possible to link our data to their test scores in Math and Spanish.[11] We match this data set with a nationwide census of teachers containing rich information on the specific subjects and classrooms taught by them every year, as well as their characteristics (e.g., gender, age, type of degree, type of contract, hours of contract). For a subset of these teachers, we recover their perceptions about the school from survey responses collected by the government. We also leverage data from a yearly school panel that includes several school characteristics such as the type of administration (e.g., public, subsidized-private, or private), an indicator if the school is located in a rural area, its total enrollment, concentration of disadvantaged students, the identity of the school principal, and whether she or he was elected through the new selection system. Finally, we complement this data with i) characteristics of the municipalities where the schools are located, ii) administrative records of complaints filed against the school, and iii) surveys of teachers run by the Ministry of Education.

Panels A and B of tables 1 and 2 present descriptive statistics of students and schools. Private schools that do not receive vouchers are excluded from our analysis. Thus, from now on, we will refer to private schools receiving vouchers either as subsidized private schools or as private schools. Preschools, adults' schools, and special education schools are also excluded. We see that students at private schools do better in terms of course grades, test scores, and grade retention. In terms of school characteristics, private schools tend to serve more students, have larger classrooms, have fewer teachers per student, and have slightly better attendance. All these differences are likely related to the fact that only a 24% of private schools are in rural areas versus a 62% of public schools. In terms of school finance, public schools receive a larger subsidy but they also serve more disadvantaged students. The share of students considered poor, and who are therefore eligible for special subsidies, is 57% in public schools and 36% in private schools.

Our analysis of the principals' labor market uses detailed administrative data from the Superintendency of Education. In Chile, every school that receives a voucher from the government must provide a detailed report of their sources of income and expenditures.

[11] As mentioned before, the SIMCE examination is only taken by students in specific grades, usually 4th, 8th, and 10th grade, and it has not been systematically run every year in the country.

These records allow us to observe all compensations paid to every school worker, by month, between 2015 and 2017. We classify compensation items into three categories: minimum wage, statutory payments, and bonuses. *Minimum wage* corresponds to a per-hour legal-minimum payment for teachers, defined by the Ministry of Education. *Statutory payments* include compensations regulated by law but unrelated to performance, such as payments for experience and for teacher certification; it also includes other compensations assigned to those who work extra hours, in rural schools, or in schools where it is "difficult" to teach according to the Ministry of Education. Finally, *bonuses* encompasses compensations related to workers' performance, such as individual and collective performance bonuses, payments from the national system of performance assessment, bonuses paid directly by the school owner in the case of private schools, and other discretionary payments and gratifications related to transportation, food, and holidays.

Since our focus is on the principals' labor market, panel C of tables 1 and 2 presents statistics of school principals' demographics and wages. On average, principals earn 2,739 USD per month. According to the representative survey CASEN 2015, the wage of school principals is placed at the 65th percentile of the wage distribution (or the 51st percentile when we only consider workers from similar cohorts and who attained higher education).[12] 42.5% of this wage corresponds to the minimum wage, 41.5% to statutory payments, and 16% to bonuses. Compared to private schools, public schools have lower wages and compensation relies more on statutory payments and less on bonuses. Finally, in terms of demographic characteristics of the school principals, public school principals are older and most of them are male; in contrast to private schools, where 61% are female.

3 Estimation of Principals' Effectiveness

In this section, we present the main model used to measure principal effectiveness and teacher value-added. We consider a specification that relates academic achievement to student characteristics, school characteristics, and to the school principal, as follows:

$$\begin{aligned} Y_{it+1} = \quad & \gamma_t + \rho_{g(i,t)} + \beta_0 A_{it} + \beta_1 f(Y_{it-1}, \bar{Y}_{it-1}, \rho_{g(i,t)}) \\ & + \theta_{p(i,t)} + \phi_0 X_{s(i,t)t} + \phi_1 \bar{X}_{s(i,t)} + \kappa_0 A_{p(i,t)} + e_{it+1}, \end{aligned} \tag{1}$$

[12]The average monthly wage of 2,739 USD corresponds to roughly 11 times the legal minimum wage.

where Y_{it+1} is a measure of student achievement, such as the course grade obtained by student i in year $t+1$, and γ_t and ρ_g stand for year and grade fixed effects. As standard in value-added models (e.g., Kane and Staiger, 2008; Chetty et al., 2014), we include a third degree polynomial at the student and classroom level in the lagged dependent variable interacted with students' grade level. We also control by student's age A_{it} and principal tenure $A_{p(i,t)}$. Since we only observe a few principals switching between schools, we do not include school fixed effects. Instead, we use correlated random effects to account for school heterogeneity (e.g., Mundlak, 1978; Chamberlain, 1980). Specifically, we include vectors of time-variant and across-time averages of the following school characteristics: total enrollment, share of disadvantaged students, share of low-income and high income parents, share of parents with a college degree, indicators for whether the school is public and for whether it is consider rural. We also include a set of controls at the level of the municipality where the school is located: average years of education, income per-capita, and the 2011 rates of crime, unemployment, and poverty.[13] Similarly to Altonji and Mansfield (2018), our approach attempts to absorb the across-schools variation in unobservable school characteristics by controlling for the school averages of its observed characteristics. Moreover, this approach helps us to overcome problems of weak identification that could arise in our case if we added school fixed effects, given that only a small fraction of principals switch schools within our time frame (e.g., Jochmans and Weidner, 2019).

In this model, principal effectiveness is captured by principal fixed effects $\theta_{p(i,t)}$. Under the standard *strict exogeneity* or *selection-on-observables* assumption, $\hat{\theta}_p$ identifies the causal impact of the principal p on student i's achievement. This parsimonious specification allows us to estimate a precise single measure of effectiveness for each principal, which may be interpreted as a weighted average effect across years if principal effects *drifted* differentially over time. Importantly, this specification does not take teacher effects into account directly, i.e., it does not include teacher fixed effects. The reason is that principals' effectiveness is at least in part mediated by teachers because one way in which principals affect student achievement is via their effects on teachers. Principals can affect teacher effectiveness through feedback and opportunities for their development, the resources they supply to teachers, and the protection they give teachers from distractions, among other mechanisms. As a result, taking teacher effects directly into account in this model would remove an important component of the variation in principal effectiveness. Nevertheless, we can use the residual from this specification to compute a relative measure of teacher effectiveness (relative to other

[13]Following Wooldridge (2010), we also add the across-time average of year fixed effects dummies to account for our unbalanced panel.

teachers working under the same school principal). Similar to most value-added models, we decompose the error term e_{it+1} into four components: teacher value-added, exogenous classroom and school level shocks, and idiosyncratic student level variation; and we use the empirical variance of these components to construct *shrunken* empirical Bayes posterior estimates of teacher effectiveness as in Kane and Staiger (2008). Appendix A presents a detailed description of the step-by-step procedure used to obtain our teacher value-added estimates.

To estimate our model, we leverage a panel at the student-subject-year level from 2011 to 2016, and we focus on the performance of the student in the two subjects for which we observe course grades every year: Math and Spanish. As mentioned before, we exclude private schools that do not receive vouchers, as well as preschools, adults' schools, and special education schools. We also exclude classes that had more than one teacher per year and eliminate the bottom and top one percent of classroom size outliers. Finally, to avoid the classification of easy-graders as high value added teachers, we restrict the sample to students for whom the teacher, in a given subject, changed between t and $t+1$. We end-up with an estimation sample of 8,687,086 student-subject-year observations, which comprises 1,935,409 students and 9,117 principals. To gauge the impact of principal effectiveness on students' learning, we estimate specification (1) again but now excluding principal fixed effects. From this, we recover a residualized measure of achievement $Y_{it+1} - \hat{Y}_{it+1}$, which we can run against our estimated fixed effects. Figure 1 visually summarizes this exercise. We find that an increase of one standard deviation in principal quality increases students' course grades by 0.17 standard deviations. This is a non-trivial effect size as reduced form evidence suggests that one standard deviation increase in GPA is associated with an 11% increase in wages, even after accounting for test score achievement and several student and school level characteristics.[14]

[14] Table C.1 in the appendix presents this result. For this exercise, we consider the universe of Chilean students who graduate from high-school in 2008 and 2009 and we look at their wages 8 to 9 years after high-school graduation.

Discussion

On the use of course grades

Our preferred measure of principal effectiveness comes from a value-added model that uses students' course grades instead of test scores as the dependent variable. The main reason for this is that students in Chile do not take nationwide standardized exams in every grade or necessarily more than once in their lives, which creates two drawbacks associated with using test scores. First, it is unclear how to compute teacher value-added without contiguous measurements of achievement, and the gap between different exams for a given student (if she takes more than one exam) is at least 4 years. Second, principals' fixed effects would be noisier as they would be estimated using a significantly smaller sample. While our model for course grades is estimated using 8,687,086 student-subject-year observations, the same model for test scores only uses 235,434 observations. Despite these issues, we estimate an alternative measure of principal effectiveness using test scores as the dependent variable of our main specification (1). As shown by table 8, all results but the one on retention of high value added teachers are robust to the use of this alternative measure of principal effectiveness.

Moreover, while it could be argued that test scores are more objective and fair than course grades, we claim that the latter is closer to what we are interested in for measuring principal effectiveness. Indeed, GPA is the metric that determines high school graduation and advancement into higher level courses,[15] and it also captures the non-cognitive dimension of students' development.[16] Furthermore, Chile has a national standardized curriculum and consequently our data gives purchase to the view that test scores and course grades can be used interchangeably as they are strongly positively correlated (e.g., Borghans et al., 2016). Figure 2 shows a strong relationship between test scores and course grades. A one standard deviation increase in course grades is associated to a 0.6 standard deviation increase in test scores. Regarding our empirical specification, we examine students' performance longitudinally, which should help to mitigate concerns about teachers or school variability in grades. Moreover, we estimate specification (1) focusing on future, instead of contemporaneous, course grades. This is particularly important to measure teacher value-added since teachers

[15] A simple linear probability model suggests that one standard deviation increase in course grades decreases the probability of grade retention by 8 pp., from a base of 4 pp.

[16] As pointed out by Bowles and Gintis (1976), grades in school not only capture cognitive skills but also perseverance, dependability, and consistency. Research has also shown that non-cognitive traits are strongly related to both school and labor market outcomes (e.g., Bowles et al., 2001; Heckman et al., 2006).

can directly impact their students' grades, potentially confounding easy-graders with high value-added teachers. In the spirit of Peteck and Pope (2019), specification (1) addresses this concern by looking at future course grades Y_{it+1} and controlling by the lagged course grades Y_{it-1} in a restricted sample with the 86% of students for whom the teacher, in a given subject, changed between t and $t+1$. Intuitively, our specification will give credit to a math teacher if her students improved their math course grade after having her as a teacher, controlling by the students' past achievement. Figure C.3 in the appendix exemplifies our approach, and table C.2 presents a comparison between the full sample and the restricted sample that we use for estimation. We do not find evidence of selective sample attrition.[17]

On the identification assumption

Despite the fact that quasi-experimental measurements of value-added seem to support the assumptions underlying observational value-added models (e.g., Kane and Staiger, 2008; Angrist et al., 2017), the use of these models as a personnel tool is still controversial (e.g., Rothstein, 2010; Chetty et al., 2014; Bacher-Hicks et al., 2014; Rothstein, 2015; Chetty et al., 2016). The main reason of concern in the teachers settings is that classroom assignments may not be exogenous conditional on the typical controls, and consequently estimates of teachers' effects based on these models should not be interpreted as causal. An analogous threat in our setting is that conditional on observables, the correlation between the assignment of students to principals depends on unobserved determinants of students' achievement. Although this threat to identification is ultimately untestable, here we discuss two points that seem reassuring.

First, we deal with the potentially endogenous sorting of students into schools using correlated random effects. Arguably, concerns about the assignment of students to principals should be ameliorated once school heterogeneity is taken into account. To asses the performance of our approach, we focus on a subset of principals who switched schools and that we observe for at least two periods in each school. For them, we estimate our main specification (1) but including school fixed effects directly. Encouragingly, this alternative measure and our main measure of principal effectiveness are strongly related. As shown by panel A of figure 3, we find that a 1 standard deviation increase in our main measure of principals'

[17] We do not find evidence of selective sample attrition in terms of grades, subject, attendance, student performance, or teachers characteristics. But, as expected, the restricted sample excludes smaller, rural, and public schools where teachers change less often from one year to another.

effectiveness implies a 0.93 standard deviation increase in our alternative measure.[18]

As a secondary check, we estimate our main specification (1) again but now including principal-school fixed effects. Then, in the spirit of a balance test, we focus on principals who moved across schools to assess how changes in a principals' effectiveness depend on the differences in predetermined school characteristics. Specifically, we estimate:

$$\hat{\theta}_{ps} - \hat{\theta}_{ps'} = \alpha + \beta(\bar{X}_s - \bar{X}_{s'}), \tag{2}$$

where $\hat{\theta}_{ps} - \hat{\theta}_{ps'}$ represents the difference in the estimated fixed effect of principal p when she worked at schools s and s' respectively; and $\bar{X}_s - \bar{X}_{s'}$ represents the difference in the predetermined characteristics (as of 2010) of schools s and s', with the vector X including: school enrollment, county enrollment, share of poor students, subsidy per student, and teachers per 100 students. Figure 3, panel B, plots the β coefficients and 95% confidence intervals obtained from separate OLS regressions of the principal effectiveness differential on the differential in predetermined characteristics of the schools. Reassuringly, we do not find evidence of an association between changes in school characteristics and changes in principal fixed effects.

In our setting, principal fixed effects would identify the causal effect of principals on students under a *strict exogeneity* or *selection on observables* assumption, i.e., conditional on observable characteristics, the correlation between the assignment of students to principals and other determinants of students achievement is innocuous. While this identification assumption is ultimately untestable —what Holland (1986) called "the fundamental problem of causal inference"— we can leverage the panel structure of our data to implement a specification diagnostic in the spirit of Rothstein (2010). Specifically, focusing on a subset of students who changed schools (and consequently principals), we can implement a falsification test. The intuition of the test is simple: if the effectiveness of the principal in the school of destination can impact students' learning in the school of origin, then the identification assumption of the model is probably wrong.

For this "validation" exercise we focus on a subset of students who switched between schools at some point and were consequently exposed to more than one principal within our time

[18] Figure C.4, in the appendix, presents a Bland-Altman plot for these measures (e.g., Bland and Altman, 1986). We do not find indications of significant bias in our main measure of principal effectiveness, relative to the alternative that includes school fixed effects.

frame. For them, we estimate the following variation of our main specification:

$$Y_{it+1} - Y_{it-1} = \alpha + \gamma_0 \hat{\theta}_p^{s_0} + \gamma_1 \hat{\theta}_{p'}^{s_1} + \phi_0 X_{s(i,t)t} + \phi_1 \bar{X}_{s(i,t)} + e_{it+1}, \tag{3}$$

where s_0 and s_1 stand for the schools of origin and destination, respectively; $\hat{\theta}_p^{s_0}$ is the estimated effectiveness of principal p who is working in the school of origin; and $\hat{\theta}_{p'}^{s_1}$ is the estimated effectiveness of principal p' who is working in the school of destination.[19] We also include schools' correlated random effects to isolate the effect of the principal from the effects of the school. The intuition of the test is simple; if the estimates suggested that a future effective principal can improve students' past achievement (i.e., $\gamma_1 > 0$), then the model could be wrong.

Table 3 presents the results from this specification diagnostic for both this model of gains and a model with lagged dependent variable. Encouragingly, we do not find much evidence of positive correlation between standardized course grade gains of the students (the pre-assignment variable) and the effectiveness of their future principal (the treatment). On one hand, the model of gains suggests a negative correlation between students' current achievement and the effectiveness of future principals. On the other hand, the model with lagged dependent variable shows a positive, but very small, effect of future principals on current achievement.[20]

4 Principal's Effectiveness and School Management

In this section, we leverage data from different sources to study the relationship between our measure of principals' effectiveness and their management practices. We evaluate if effective principals can impact the retention of teachers and the sorting of students across classrooms, and whether they are recognized by teachers and the school community.

[19]For this exercise we consider a "jackknife" estimate of principal effectiveness, i.e., to avoid spurious correlation, we estimate principal effectiveness excluding students who switched schools.

[20]While this is consistent with the idea that models with lagged dependent variable do not really reflect "value-added", our results are in line with Chetty et al. (2017), as the effect size on future principal effectiveness is quite small compared with the main effect, which remains stable.

Management of workforce: teachers' exit rate

Teacher shortages and high turnover rates have recently received considerable attention from policymakers, as they impose financial costs on schools and may affect students outcomes. Teacher turnover is at the center of this debate, given the well-known difficulties experienced by new teachers, and the fact that teacher turnover seems to have adverse effects on school effectiveness, specially in lower-achievement schools (e.g., Ronfeldt et al., 2013; Hanushek et al., 2016). Motivated by the known negative correlation between wages and labor turnover (e.g., Manning, 2011), researchers have attempted to estimate the turnover elasticity in the labor market of teachers, obtaining different results. For instance, Hanushek et al. (2004) finds no effects of wages on the probability that teachers leave the profession, while Ransom and Sims (2010) find a highly significant effect. Closer to the latter, Hendricks (2014) estimates an average turnover elasticity of -1.4 and reports that it is larger for inexperienced teachers.

In the spirit of previous work, we study the relationship between teacher turnover, wages, and other school characteristics, and we present novel evidence on the effect of principal quality on teachers' retention. We leverage a rich panel of teachers, including teacher and school characteristics from 2008 to 2016, to estimate the following linear probability model:

$$\text{Exit}_{jt+1} = \gamma_t + \rho_{m(j,t)} + \beta\hat{\theta}_{p(j,t)} + \Phi_1 X_{jt} + \Phi_2 X_{s(j,t)t} + \epsilon_{jt+1}, \tag{4}$$

where Exit_{jt+1} equals 1 if teacher j experiences either a job-to-job or a job-to-non-teaching employment transition the next period $t+1$, while γ_t and $\rho_{m(j,t)}$ stand for year and municipality fixed effects. We also include controls at the school and teacher level. Specifically, the vector $X_{s(j,t)t}$ controls for school-level characteristics such as the average wage paid to teachers, the share of poor students, and the school size. We also include type of school dummies indicating whether the school is public or not, and whether it is rural or not. The vector X_{jt} has teacher-level controls including gender, age, age squared, and an indicator if the teacher holds a permanent (instead of a fixed-term) contract. The parameter of interest is β and it reflects the association between the standardized measure of principal effectiveness obtained in section 3 and teachers' exit rate.

Column 1 of table 4 shows the main result. We find that a 10% increase in wages decreases the likelihood that a teacher leaves the school by 6%, implying an observational turnover elasticity of around -0.6, half the one reported in Hendricks (2014). Moreover, a one standard deviation increase in principal effectiveness decreases the average exit rate of teachers by 3%. To explore heterogeneous effects by teachers' experience and effectiveness, we estimate model (4) separately for teachers of different ages and of different estimated value-added. Table 4, columns 2 to 4, present the analysis by teachers' age. We start by noticing that the average exit rate decreases monotonically with the age group, from 15% among the youngest teachers to 5% among the oldest. Our results suggest that principal effectiveness is negatively associated with the turnover of teachers between 22 and 37 years old. A one standard deviation increase in principal quality decreases the likelihood that these young teachers will leave the school by 4%, but it has no statistically significant effects on the exit rate of the oldest groups of teachers. Regarding teacher quality, columns 5 to 7 present the results for different groups of teachers classified based on their location within the empirical distribution of our estimated measure of teacher value-added.[21] Results suggest that a one standard deviation increase in principal quality decreases teacher turnover of medium and high value-added teachers by 3% and 5%, respectively. However, the effect is not different from zero for low value-added teachers.[22]

Interestingly, table 4 also shows that the relationship between principal effectiveness and teacher turnover is mostly driven by private schools. If principals can recognize good teachers, as our analysis suggests and previous research has shown (e.g., Jacob and Lefgren, 2008), then giving principals more discretion in the management of teachers could be an effective way to use that local knowledge, especially in public schools where bureaucratization may hinder the use of it. Our results also suggest that policies aimed at hiring better principals may be a cost-effective approach to decrease teacher shortages and increase teacher quality. According to our estimates, the average *willingness to pay* of teachers for a one standard deviation more effective principal is equivalent to a 5% increase in their wages. Thus, replacing the bottom 5% of principals by the average principal could save around 43,200 USD per school-year, the equivalent of 18 yearly vouchers for students.[23]

[21]Notice that teacher value-added is estimated from the residual of equation (1), as described in Appendix A. Consequently, our measure of teacher value-added is relative to the school principal, i.e., it is computed from deviation in the achievement of different classrooms exposed to the same school principal.

[22]All these results are robust to non-linear transformations of model (4), i.e., a logistic regression model. Moreover, as shown by table C.3 in the appendix, most results but the one on the retention of young teachers hold if we add teacher fixed effects to specification (4), i.e., to leverage within-teacher variation in principal effectiveness.

[23]The average monthly wage paid to teachers in Chile is 1,500 USD, and schools have 24 teachers on average. The average annual subsidy per student corresponds to 2,424 USD.

Management of classrooms: students' sorting

Principals could also affect schools through the assignment of students to classrooms (tracking) and the assignment of classrooms to teachers (matching). A rationale for this could be to help teachers to target their lessons (e.g., Duflo et al., 2011; Card and Giuliano, 2016). As pointed out by Rothstein (2010), sorting may also happen through efforts on the principal's part to reward favored teachers through the allocation of easy-to-teach students, or through parental requests (e.g., Jacob and Lefgren, 2007; Kalogrides and Loeb, 2013). Thus, the assessment of tracking is also important as it represents a potential source of bias in models of teacher value-added (e.g., Horváth, 2015). To assess the extent of tracking in Chile, and to explore how principal's effectiveness is related to it, we construct a sorting index at the school-year level à la Kremer and Maskin (1996). Specifically, we measure how students sort across classrooms by estimating the following regression:

$$\bar{y}_{cg} = \alpha_g + \tau_{st} y_{i(c,g)} + \epsilon_{cg}, \tag{5}$$

where $\bar{y}_{cg}$ stands for the average course grade of classroom c of grade g, and $y_{i(c)}$ represents the course grade of student i who sits in classroom c. We estimate this specification by school and year, including grade fixed effects. The sorting index is given by our estimate of τ_{st}:

$$\tau_{st} = \frac{Cov(\bar{y}_{cg}, y_{i(c,g)})}{Var(y_{i(c,g)})}.$$

Intuitively, perfect sorting is a case in which all variation in $y_{i(c,g)}$ comes from variation between instead of within classrooms. Indeed, this metric is equivalent to the R^2 from a regression of $y_{i(c,g)}$ on classroom dummies. We first evaluate the prevalence of students' sorting in Chilean schools. Figure 4, Panel A, plots the histogram of our estimated sorting index τ_{st}. We find that, on average, $\hat{\tau}_{st}$ is 0.049 with a standard deviation of 0.041. In order to benchmark this number, we construct a theoretical upper bound by ranking students within a school-grade and sorting them across all classrooms in a given grade in the order determined by their ranking. The average upper bound that we obtain from this procedure is 0.690 with a standard deviation of 0.079. From this exercise, we conclude that sorting based on course grades is not very common in Chile as it only represents a 7% of this empirical upper bound.

In spite of the low prevalence of sorting in Chilean schools, we explore the association between principal effectiveness and our sorting index. Figure 4, Panel B, depicts a bin scatter with the relationship between the sorting index and principal effectiveness in public and private schools. We find that a one standard deviation increase in principal effectiveness increases sorting by 0.005 pp over a base of 0.053 in public schools and decreases it by 0.002 pp over a base of 0.046 in private schools. Although these effect sizes may seem large, the overall levels of sorting are still small.

Managerial practices: Teachers' perceptions and complaints

Similar to recent research looking at managerial practices within schools (e.g., Di Liberto et al., 2015; Bloom et al., 2015), in this subsection we explore how teachers perceive the practices of their principals. Specifically, we analyze a set of surveys, available for 2010, 2011, 2014, and 2015. These surveys ask teachers about their level of agreement with different statements, such as *The principal does a good job* and *The principal promotes a good work climate*. Every teacher must provide an answer within a range from 1 to 4 (or from 1 to 5 in some years), where 1 represents high disagreement with the statement and 4 (or 5) represents a high level of agreement. We use their responses to create a dummy variable at the survey respondent level that equals one if the teacher "highly agrees" with the statement about the principal, i.e., her response is at the top of the specific scale for that question. Then, we take the average across respondents at the school-year level and assign this to the corresponding school principal. Using this principal-level data set, we estimate the following model:

$$y_p = \phi_{m(p)} + \beta\hat{\theta}_p + \epsilon_p, \tag{6}$$

where y_p represents the fraction of respondents working with principal p that "highly agree" with statement y, $\phi_{m(p)}$ is a municipality fixed effect, and $\hat{\theta}_p$ is the standardized measure of principal effectiveness discussed in section 3. We present our results in panel A of table 5. Column 1 presents the estimated β coefficients and column 2 presents bootstrap standard errors (100 replications) to account for the fact that $\hat{\theta}_p$ is an estimate itself. Column 3 presents the p-values adjusted for multiple testing using the step-down procedure proposed in Romano and Wolf (2005), which controls by the family-wise error rate to prevent us from declaring a true null hypothesis to be false.[24] Moreover, column 4 reports a one-sided p-

[24]A similar approach is taken by Bassi et al. (2016) in the context of heterogeneous treatment effects of a policy evaluation in Chile.

value obtained from a permutation test. To compute this number, we randomly reshuffled our principal fixed effects 1,000 times, and then we calculate the proportion of sampled permutations where the absolute value of the coefficients obtained using the reshuffled fixed effect was greater than or equal to our $\hat{\beta}$ estimate. Intuitively, this exercise gauges how likely would it be to obtain our results just by chance. Finally, columns 5 and 6 present the mean of the dependent variable and the number of observations used for estimation. The number of observations changes because not all questions are asked every year.

We find that more effective principals do a better job according to teachers. Results in Table 5 show that more effective principals are associated with a larger fraction of their teachers highly agreeing with positive statements about their management. Ordered by effect size, we find that one standard deviation increase in principal effectiveness increases agreement with the statements *principal does a good job* by a 5%; *principal promotes a good work climate*, *principal knows students needs*, *principal is good at communicating*, and *principal is a good manager* by 4%; *principal engages parents*, *principal engages teachers*, and *principal knows teacher needs* by 3%; and *principal makes good decisions*, *principal can be trusted*, *principal includes teachers*, and *principal is effective* by 2%. Reassuringly, all these results are robust to our permutation test and to adjustments in our inference strategy to control for multiple testing.

Finally, we leverage administrative data of complaints filed by the school community against the school. This data is recorded to monitor that holders of educational establishments comply with the laws, regulations, and instructions issued by the Superintendency of Education. As before, we build a data set at the principal level. Specifically, we consider the number of complaints filed against the school per 100 students and we sum them across years to then assign them to a principal. Then, we estimate specification (6) but using the average number of complaints per 100 students as the dependent variable y_p. Table 5, panel B presents our results. We find that a one standard deviation increase in principal effectiveness is associated with a decrease in the number of complaints filed against a school. The most important and robust effects correspond to a decrease in complaints for bullying and denied enrollment. Complaints for the former decrease by 17%, a result that we interpret as indicative of an association between principal effectiveness and more discipline within the schools, while complaints for the latter decrease by 18%, suggesting that principal effectiveness is not related to a cream-skim of students. There is also suggestive evidence that better principals decrease complaints related to the absence of teachers and bad school infrastructure by 18%

and 36%, although these results do not pass our additional robustness checks.[25]

The evidence presented here suggests that better principals are recognized by the school community. However, is the principal's effectiveness also recognized in the labor market?

5 Principal's Effectiveness and Wages

We use administrative data from public and subsidized private schools to study the relationship between our measure of principal effectiveness and principals' compensation. In particular, we estimate the following regression model:

$$\ln(\text{wage}_{pt}) = \alpha + \rho_{m(p,t)} + \gamma_t + \beta\hat{\theta}_p + \phi\boldsymbol{X}_{pt} + \epsilon_{pt}, \quad (7)$$

where $\ln(\text{wage}_{pt})$ represents the logarithm of the average hourly wage paid to principal p at time t, γ_t are year fixed effects, and $\rho_{m(p,t)}$ is a fixed effect at the level of the municipality in which principal p works at time t. The parameter of interest is β, and it represents the association between a principal's wage and her standardized effectiveness $\hat{\theta}_p$. The vector $\boldsymbol{X}_{pt}$ includes principal level controls such as tenure, tenure squared, an indicator for whether the principal is female, and for whether she has a permanent contract.

Public sector compensation usually does not include pay for performance (Finan et al., 2015), and although there is a good rationale for this,[26] it has been argued that fixed compensation schemes make it difficult to attract and keep the best personnel in public schools. This discussion, which has motivated several studies on the effects of pay for performance (e.g., Rothstein, 2015; Cullen et al., 2016; Biasi, 2018) and teacher firing policies (e.g., Staiger and Rockoff, 2010; Boyd et al., 2011; Cowen and Winters, 2013), is also relevant to the Chilean

[25] We complement these results by estimating the impact of principal effectiveness on the average agreement of teachers across statements and on the total number of complaints filed against the school. Then, we check the robustness of these relationships to adding randomly-selected subsets of school-control variables, as in Card et al. (2018). Figure C.5 in the appendix shows the results from this exercise. We observe that, for any number of control variables, the average and the median point estimate across randomizations is very close to our baseline estimate for both: the mean agreement of teachers and the total number of complaints.

[26] Performance pay for bureaucrats can create severe multi-tasking problems, where bureaucrats focus on the incentivized dimension of their job at the expense of the non-incentivized dimension (Holmstrom and Milgrom, 1987). While multi-tasking is an issue in many contexts, it can be particularly severe in public sector contexts where agents wield substantial authority (e.g., police, judges, public school principals) and in which it is hard to find an objective measure of the *truth* on which to incentivize them.

case. Figure C.6, panel A plots the distribution of hourly wages, residualized with respect to year and county fixed effects, in both private and public schools. Perhaps not surprisingly, we find that public school wages are significantly less spread and are 0.09 log points lower than in the private schooling sector. Moreover, panel B of figure C.6 shows that, like in the US, wages in public schools rely less on pay-for-performance. On average, the bonus component of wages represents 22% of the principal's salary in private schools but only 9% in public schools. To analyze how different components of wages relate to principal effectiveness, we decompose the $\ln(\text{wage}_{pt})$ into $\ln(\text{base}_{pt})$ and $\ln(\text{wage}_{pt}/\text{base}_{pt})$, where "base" corresponds to the sum of the minimum legal wage and the statutory payments described in section 2, i.e., base_{pt} corresponds to the total wage minus the bonuses.

The estimation results from these regressions are presented separately by private and public schools in table 6. Panel A shows the results for $\ln(\text{wage}_{pt})$ with and without controls; and panel B shows the results (with controls) for the two dependent variables $\ln(\text{base}_{pt})$ and $\ln(\text{wage}_{pt}/\text{base}_{pt})$. We find that a one standard deviation increase in principal effectiveness is associated with a 4% increase in wages in private schools, but it is related to only a 1% increase in wages in public schools. We also find that the tenure profile is concave and more relevant to explain wage variation in public schools, and that women earn lower wages, especially in private schools. Regarding the association between principal effectiveness and wages components, panel B shows that in private schools, about half of the association between principal effectiveness and wages is accounted for by the payments over the base, and this figure rises to about two-thirds in public schools.

In spite of the fact that both sectors seem to reward principal effectiveness through bonuses (as shown by panel B), the higher "quality premium" paid by private schools (shown in panel A) would suggest that public schools are worse equipped to attract high-quality principals. Motivated by these results, the next section presents a simple two-sided matching model to guide our thinking and empirical inquiry on how public schools can attract effective principals.

6 A Two-Sided Matching Model

This section builds on Logan (1996) to simultaneously investigate schools' preferences to offer a job and workers' choice given the job offers. The model is based on an underlying random matching model of the labor market, which itself is a stochastic variant of deterministic two-sided matching models studied in game theory (e.g., Roth and Sotomayor, 1990).[27] The timing of the model is the following:

- Workers apply to all available schools.
- Schools evaluate applicants and make offers according to a decision rule.
- Workers evaluate the received offers and choose the highest-utility alternative.

The school's decision

Similar to Abowd and Farber (1982), an underlying random utility model is defined to describe the decision of schools regarding whether or not to make jobs available to particular workers. For school j, the utility of hiring worker i of ability θ_i is defined as:

$$U_j(i) = m_j + \beta_j \theta_i + \epsilon_{1ij}, \tag{8}$$

while j's utility of not hiring worker i is:

$$U_j(\neg i) = s_j + \epsilon_{0ij}, \tag{9}$$

where m_j represents market effects on the utility of hires in general (e.g., reflecting the need for filling the position), β_j is the increase in utility that the school would experience from hiring a worker of marginally higher quality, and s_j is simple a baseline utility that school j derives from its present state of staffing. Finally, ϵ_{1ij} and ϵ_{0ij} represent factors that are not known to the observer but that influence the utility of school j of hiring or not hiring worker i.

When expression (8) is greater in value than expression (9), employer j makes a job available: $o_{ij} = 1$, zero otherwise. Thus, the exact probability that school j will make an offer depends

[27] This game is a random variant of the "college admissions" game of the formal game theory literature, and because the deterministic results are transferable to the random matching game, it is known that at least one stable matching of employers and workers exists such that no worker-employer pair who are not matched to each other can improve their utilities by abandoning any current pair and establishing a new match together.

on the distribution of the differences between the two error terms, as well as on the non-stochastic parts of (8) and (9). If ϵ_{1ij} and ϵ_{0ij} are *iid* type I extreme value, then the difference will follow a logistic distribution, and the probability that j will make an offer is given by:

$$Pr(o_{ij}) = \frac{\exp(\beta_{0j} + \beta_j \theta_i)}{1 + \exp(\beta_{0j} + \beta_j \theta_i)}, \tag{10}$$

where $\beta_{0j} = m_j - sj$, and the offer of unemployment is always available to the workers, i.e., $Pr(o_{i0}) = 1$.

The worker's decision

Assuming that employers act independently of one another, conditional on workers' quality θ_i, then each applicant would be presented some set O_k of offers from the employers as a whole. There will be $R = 2^J$ distinct possible offering sets when J employers make separate decisions. Given this, the probability that worker i obtain a given offering set O_k is given by:

$$Pr(S_{ik}) = \prod_{m \in O_k} Pr(O_{im} = 1) \prod_{n \in \bar{O}_k} Pr(O_{in} = 0), \tag{11}$$

where m is an element (offer) of set K and n is an element of the complement set of O_k. A worker will choose her most preferred offer from the offering set that she faces. This is specified as a second random utility model. The indirect utility that worker i obtains from the job offered by employer j is defined as:

$$V_{i(j)} = h_j + w_j \theta_i + v_{ij}, \tag{12}$$

where h_j represents a baseline level of payments and amenities, w_j is a pay-for-performance component offered by the employer, and v_{ij} represents idiosyncratic preferences of the worker for a given job. Workers evaluate simultaneously every job offer that they find available to choose the one that delivers the highest utility. If v_{ij} follows a type I extreme value distribution, then the probability that worker i selects job j given the set of offers O_k is given by this polytomous conditional logit:

$$Pr(A_{ij} \mid O_k) = \begin{cases} \frac{\exp(h_j + w_j \theta_i)}{\sum_{h \in O_k} \exp(h_h + w_h \theta_i)} & , \quad j \in O_k \\ 0 & , \quad j \notin O_k. \end{cases} \tag{13}$$

Given our assumptions about the distribution of the random components in (8), (9), and (12), and further assuming that these random components are mutually independent, the probability that worker i ends-up in job j is given by:

$$\begin{aligned} Pr(A_{ij}) &= \sum_{k=1}^{R} Pr(A_{ij} \mid S_{ik}) \times Pr(S_{ik}) \\ &= \sum_{k=1}^{R} Pr(A_{ij} \mid S_{ik}) \times \prod_{m \in O_k} Pr(O_{im} = 1) \times \prod_{n \in \bar{O}_k} Pr(O_{in} = 0) \\ &= \sum_{k: j \in O_k} \frac{\exp(h_j + w_j \theta_i)}{\sum_{h \in O_k} \exp(h_h + w_h \theta_i)} \times \prod_{m \in O_k} \frac{\exp(\beta_{0m} + \beta_m \theta_i)}{1 + \exp(\beta_{0m} + \beta_m \theta_i)} \\ &\quad \times \prod_{n \in \bar{O}_k} \frac{1}{1 + \exp(\beta_{0n} + \beta_n \theta_i)}. \end{aligned}$$

Importantly, from this model we can obtain the expected quality of the workforce in a given school, which depends on the choices of both sides of the labor market. The expected quality of the workforce in school j is given by:

$$\mathrm{E}[\theta_i \mid \text{school} = j] = \int_{\theta} \theta_i f_{\theta \mid \text{school}=j}(\theta_i \mid \text{school} = j) d\theta.$$

Simulation

We are interested in the allocation of worker quality in the public and private sectors. More specifically, we seek to understand how the allocation of principal effectiveness in a given sector depends on the *selection* parameter β_j and the *pay-for-performance* parameter w_j of the model. For this purpose, we will consider a particular case of the model with only two schools, one private and one public. In this setting, there are only four possible offering configurations from public and private schools $(p, v) \in \{(0,0), (0,1), (1,0), (1,1)\}$. Thus, the probability that worker i is at a public school given her quality is given by:

$$Pr(A_{ip} \mid \theta_i) = \Bigg(\frac{\exp(h_p + w_p\theta_i)}{\exp(h_p + w_p\theta_i) + \exp(h_v + w_v\theta_i)} \times \frac{\exp(\beta_{0p} + \beta_p\theta_i)}{1 + \exp(\beta_{0p} + \beta_p\theta_i)} \times \frac{\exp(\beta_{0v} + \beta_v\theta_i)}{1 + \exp(\beta_{0v} + \beta_v\theta_i)} \Bigg) + \Bigg(1 \times \frac{\exp(\beta_{0p} + \beta_p\theta_i)}{1 + \exp(\beta_{0p} + \beta_p\theta_i)} \times \frac{1}{1 + \exp(\beta_{0v} + \beta_v\theta_i)} \Bigg). \tag{14}$$

In this case, the expected principal effectiveness in the public school is given by:

$$\mathrm{E}[\theta_i \mid \text{Public}] = \int_\theta \theta_i f_{\theta|\text{Public}}(\theta_i \mid \text{Public}) d\theta \tag{15}$$

From Bayes' rule, we know that:

$$f_{\theta|p}(\theta_i \mid \text{Public}) = \frac{Pr(A_{ip} \mid \theta_i) \times f_\theta(\theta_i)}{Pr(\text{Public})},$$

where $Pr(A_{ip} \mid \theta_i)$ is given by (14) and $Pr(\text{Public})$ is a scale factor equal to the fraction of public schools (0.5 in this case). Assuming that $f_\theta(\theta_i)$ is a standard normal, we can compute $\mathrm{E}[\theta_i \mid \text{Public}]$ using numerical integration. More importantly, we can study how this object depends on β_p and w_p, two relevant parameters related selection and payment policies in public schools.

Figure 5 shows how the allocation of principal effectiveness in public schools depends on the selection and payment parameters. To construct this figure, we created a grid for β_p and w_p from 1 to 10, and we compute (15) for each cell of this grid. We set all other parameters $(h_p, h_v, \beta_{0p}, \beta_{0v}, w_v, \beta_v)$ at constant baseline levels (1,1,1,1,5,5). As shown by this simulation, the payment component w generates Roy (1951) selection based on workers' quality θ_i. The intuition is that higher quality candidates would benefit more from a larger share of payments related to performance.[28] However, as the model and simulation highlight, the association between wages and quality is not the only driver of choice. Workers' choices are

[28] This intuition does not necessarily hold if workers are risk averse and the payment is based on a noisy signal of workers' quality. While this point is important, the main insight of our model is on workers' selection and not compensation, and consequently we leave this extension for future work.

constrained by the offers that they receive, which in turn depend on the employer selection parameter. Thus, personnel selection can accentuate or counteract the sorting of workers based on the payment component. Moreover, poor selection policies may lead to a loss of talented and motivated workers who would take a position despite a low payment because of strong idiosyncratic preferences.[29] In what follows, we leverage quasi-experimental variation to estimate the effects that a new personnel selection policy had on public schools.

7 Selection of School Principals

With the purpose of improving the management and administration of public education and of strengthening managerial performance, in 2011 a new system to select public school principals was established. Before the reform, the appointment of public school principals was the exclusive responsibility of the municipalities; this process was unsupervised by the central government and was therefore more prone to patronage (e.g., Akhtari et al., 2018; Colonnelli et al., 2019). Since the reform, principals are elected through public, competitive, and transparent competitions in a process that is lead by a third-party human resources firm and that is overseen by the Civil Service, an agency responsible for selecting the highest-level bureaucrats of the central government. As established by Law, the Civil Service is mandated to act as "the guarantor of the merit and suitability of the applicants, in public, competitive, and transparent competitions to recruit professionals with pedagogical leadership, management capacity, and strategic vision".

The implementation of the new selection system was part of a more comprehensive reform intended to level conditions between public and private schools (Law 20.501). Since this reform was enacted, public school principals are allowed to form their own management teams (i.e., the Deputy Director, the Inspector General and the Chief Technician of the school) without having to call a contest for those position and they can fire up to 5% of teachers with a bad or regular teacher evaluation. Moreover, in recognition of the relevance of principals, the reform also increased their compensation in accordance with a deterministic rule based on the number of students enrolled and the concentration of disadvantaged students in the establishment in which they work.[30]

[29]Idiosyncratic preferences may come, for instance, from pro-social behavior (e.g., Dal Bó et al., 2013; Ashraf et al., 2018; Deserranno, 2019) or preferences for payment stability (e.g., Rothstein, 2015).

[30]We leverage this arbitrary rule to study the effect of higher wages on principal effectiveness. This analysis is presented in Appendix B. Although suggestive, evidence on the validity of the research design

In this section, we leverage the timing of the adoption of this new selection system to study its impact on the allocation of principal effectiveness. We use digitized data from all the competitions for principal positions between 2012 and 2016 to identify the time when a school opened a vacancy and the time when a new principal was appointed. As shown by figure C.7 in the appendix, the adoption of this system was staggered. This is because the replacement of principals was not mandatory, and contests did not always succeed at finding a principal. As expected, the number of principals elected under the new regime increased over time, with around 370 new principals elected every year since 2012. Table C.4 in the appendix presents some descriptive statistics to compare schools based on their adoption of the new system. Columns 1 and 2 present the average and standard deviation of school characteristics for schools that did not adopt the new system and those that adopted it within our time frame. Overall, we observe that schools that adopted the new selection system tend to be larger, less rural, have fewer poor students, and have better test scores than non-adopters. Consistently, these schools are located in municipalities with higher income per-capita and more years of schooling. In columns 4 and 5 we present the characteristics of early (2012 and 2013) versus late (post 2014) adopters. We find that early adopters tend to be larger and less rural, but differences are smaller, and we find no differences in tests scores. Moreover, there are no clear differences in terms of municipality characteristics.

Regarding principals' characteristics, in table C.5 we present the differences between principals who were selected by the new system and those who were not. We find that principals selected under the new system are less likely to have worked as teachers and are more likely to have worked at administrative positions (e.g., Deputy Director, Inspector General, etc.). Interestingly, they are also more likely to have worked in the private sector in the past. Regarding their characteristics, principals elected with the new system are younger, more likely to have a college degree, and slightly more likely to be a female. In the same vein of the previous descriptive statistics, table C.6 compares the characteristics of the schools of origin with those of the school of destination for principals elected with the new system. We find no differences in terms of school wages, but it seems that these principals are leaving schools with more students, in rural areas, and with fewer resources per student (a byproduct of having fewer disadvantaged students for whom there is a targeted voucher). In line with a story of amenities, these principals are arriving in municipalities that on average have higher income and more years of schooling.

(RDD) is not conclusive, and that led us to exclude this exercise from the main results.

Empirical strategy and results

To assess the effects of this new selection system, we estimate the following staggered difference-in-differences regression:

$$y_{st} = \alpha_s + \alpha_t + \beta ADP_{st} + \sum_t \Phi_t' X_s I[year = t] + \epsilon_{st}, \tag{16}$$

where s and t stand for school and year, and the dependent variable y_{st} corresponds to the standardized version of our measure of principal effectiveness discussed in section 3. In Chile, the new selection system is known as ADP (the acronym of *Alta Dirección Pública*), thus ADP_{st} is a dummy variable that takes the value one from the first year the school selected a principal using this system, and X_s is a vector of predetermined school and municipality characteristics (measured in 2010).[31] Finally, α_s and α_t are school and year fixed effects, and ϵ_{st} is an error term robust to heteroscedasticity.

Our parameter of interest is β, and it captures the change in principal effectiveness after a school selects its principal using the new system. Results are shown in Table 7. In light of Athey and Imbens (2018), we report robust standard errors in parenthesis and robust standard errors clustered at the school-level in brackets. Estimates from column 1 suggest that after a school selects a principal using the ADP system, there is a statistically significant increase in principal effectiveness of 0.043 standard deviations. Columns 2 to 4 show that controlling flexibly by school and municipality characteristics during the pre-reform period does not affect the significance nor the effect size of our estimates. In column 5, we follow Crump et al. (2009) and truncate our analysis sample based on a propensity score that estimates the probability that a school selects a principal under the new system. We estimate an effect of 0.048 standard deviations using this truncated sample. Column 6 presents the results when we keep only those schools that adopted the new system, and therefore we identify the effect of ADP selection from the timing of adoption among adopters. We find slightly smaller but overall similar results: schools that selected a principal using the ADP system earlier experienced a statistically significant increase in principal effectiveness of 0.042 standard deviations. Finally, as a robustness check, we perform two placebo exercises. First, we consider any principal turnover that happened before the reform (2008-2011) in public

[31] The set of school controls includes income per student, share of disadvantaged students, total enrollment, and 4th grade Math and Spanish test scores. Municipality level controls include poverty rate, average household income, unemployment rate, average years of education, and literacy rate.

schools as a treatment. Second, we consider any principal turnover that happened after the reform (post 2011) in private schools as a treatment. In both cases, the placebo treatment takes the value one if there is a change in the principal and stays as a one afterwards. Reassuringly, we find that turnover itself does not impact principal effectiveness in these placebo exercises.

The key identification concern in our setting is that conditional on time-invariant school characteristics, year aggregate shocks, and differential trends parametrized by pre-reform school and municipality characteristics, there might still be unobserved confounding factors that correlated with the timing of adoption of this new system and other determinants of principal quality. To partially address this concern, we estimate a variation of model (16) with a dynamic treatment. In particular we estimate the following event-study type of regression:

$$y_{st} = \alpha_s + \alpha_t + \sum_{j=-4}^{-2} \beta_j I[k = j] + \sum_{j=0}^{4} \beta_j I[k = j] + \sum_{t} \gamma_t' X_s I[year = t] + \epsilon_{st}, \qquad (17)$$

where k is the year relative to the first year the school selected a principal with the new system. The estimation sample includes all public schools independent of whether they elected a principal via the new system or not. For those schools that used the new system, we include a window of four years around the adoption to facilitate the study of the timing of the effect. Our results are robust to not including this window restriction.

Figure 6 presents this dynamic version of the previous staggered difference-in-differences estimation. This figure provides a visual support for our identification strategy, as we observe parallel trends in the pre-period. In the pre-period, the estimates tend to be negative but small, around zero, and not significant. The slightly negative effect in the pre-period may suggest that schools that were going to appoint an ADP principal had some turnover leading to lower effectiveness of the principal in the pre-period. Nevertheless, the effect size on principal effectiveness after her selection via the ADP system suggests that our results are not reflecting reversion to the mean. Indeed, in the after period we observe an increase in principal effectiveness that remains stable over time. As shown by Panel B, we find similar patterns when we flexibly control for pre-reform school characteristics.

8 Conclusion

In this paper, we present a novel extension of the canonical teacher value-added model. Our estimation leverages large administrative data sets to measure principals' and teachers' effectiveness while accounting for schools' heterogeneity. We find substantial variation across principals in their ability to improve students' learning. A one standard deviation increase in principal effectiveness raises students' course grades and test scores by 0.17 and 0.24 standard deviations, respectively. This finding contributes new evidence on the role that school managers play in improving students' achievement. Leveraging our measure of principal effectiveness, we show that better principals can decrease teacher turnover by increasing the retention of young and high value-added teachers. In line with this, we found suggestive evidence that teachers and the school community can recognize effective management. Thus, policies aimed at hiring better principals may be a cost effective approach to decrease teacher shortages and improve school outcomes. Reassuringly, we also find that principals' effectiveness is rewarded by the labor market, especially by private schools.

The smaller return to effectiveness in the public sector may be a hurdle to attract effective leaders into the schools that most need them. However, a two-sided selection model shows that one mechanism through which public schools could attract high-quality candidates with strong preferences for public schools is by improving personnel selection. In 2011, Chile implemented a reform to level conditions between public and private schools, with a special focus devoted to the improvement of leadership. Since then, public school principals are elected through public, competitive, and transparent competitions in a process that is lead by a third-party human resources firm and that is overseen by the Civil Service of the central government. Using quasi-experimental variation in the timing of adoption of this new selection system, we show that it increases the average quality of principals in public schools by 0.04 standard deviations, which is enough to close one-half of the public-private achievement gap in about 6 years. How public schools fare in the long run as a consequence of this reform is an important task for future work.

References

Abowd, J. M. and H. S. Farber (1982): "Job queues and the union status of workers," *ILR Review*, 35, 354–367.

Akhtari, M., D. Moreira, and L. Trucco (2018): "Political turnover, bureaucratic turnover, and the quality of public services," *Working Paper*.

Altonji, J. G. and R. K. Mansfield (2018): "Estimating group effects using averages of observables to control for sorting on unobservables: School and neighborhood effects," *American Economic Review*, 108, 2902–46.

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters (2017): "Leveraging lotteries for school value-added: Testing and estimation," *The Quarterly Journal of Economics*, 132, 871–919.

Ashraf, N., O. Bandiera, and S. Lee (2018): "Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services," .

Athey, S. and G. W. Imbens (2018): "Design-based analysis in difference-in-differences settings with staggered adoption," Tech. rep., National Bureau of Economic Research.

Bacher-Hicks, A., T. J. Kane, and D. O. Staiger (2014): "Validating teacher effect estimates using changes in teacher assignments in Los Angeles," Tech. rep., National Bureau of Economic Research.

Bassi, M., C. Meghir, and A. Reynoso (2016): "Education quality and teaching practices," Tech. rep., National Bureau of Economic Research.

Bender, S., N. Bloom, D. Card, J. Van Reenen, and S. Wolter (2018): "Management practices, workforce selection, and productivity," *Journal of Labor Economics*, 36, S371–S409.

Béteille, T., D. Kalogrides, and S. Loeb (2012): "Stepping stones: Principal career paths and school outcomes," *Social Science Research*, 41, 904–919.

Biasi, B. (2018): "The Labor Market for Teachers Under Different Pay Schemes," Tech. rep., National Bureau of Economic Research.

Bland, J. M. and D. Altman (1986): "Statistical methods for assessing agreement between two methods of clinical measurement," *The lancet*, 327, 307–310.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013): "Does management matter? Evidence from India," *The Quarterly Journal of Economics*, 128, 1–51.

Bloom, N., C. Propper, S. Seiler, and J. Van Reenen (2015): "The impact of competition on management quality: evidence from public hospitals," *The Review of Economic Studies*, 82, 457–489.

Bloom, N. and J. Van Reenen (2007): "Measuring and explaining management practices across firms and countries," *The quarterly journal of Economics*, 122, 1351–1408.

Borghans, L., B. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2016): "What grades and achievement tests measure," *Proceedings of the National Academy of Sciences*, 113, 13354–13359.

Bowles, S. and H. Gintis (1976): "Schooling in Capitalist America," *New York: Basic.*

Bowles, S., H. Gintis, and M. Osborne (2001): "The determinants of earnings: A behavioral approach," *Journal of economic literature*, 39, 1137–1176.

Boyd, D., H. Lankford, S. Loeb, M. Ronfeldt, and J. Wyckoff (2011): "The effect of school neighborhoods on teachers' career decisions," *Whither opportunity*, 377–396.

Branch, G. F., E. A. Hanushek, and S. G. Rivkin (2012): "Estimating the effect of leaders on public sector productivity: The case of school principals," Tech. rep., National Bureau of Economic Research.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82, 2295–2326.

Canay, I. A. and V. Kamat (2017): "Approximate permutation tests and induced order statistics in the regression discontinuity design," *The Review of Economic Studies*, 85, 1577–1608.

Card, D., A. Fenizia, and D. Silver (2018): "The health effects of cesarean delivery for low-risk first births," Tech. rep., National Bureau of Economic Research.

Card, D. and L. Giuliano (2016): "Can tracking raise the test scores of high-ability minority students?" *American Economic Review*, 106, 2783–2816.

Cattaneo, M. D., M. Jansson, and X. Ma (2018): "Manipulation testing based on density discontinuity," *The Stata Journal*, 18, 234–261.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 47, 225–238.

Chetty, R., J. Friedman, and J. Rockoff (2014): "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American Economic Review*, 104, 2593–2632.

——— (2016): "Using lagged outcomes to evaluate bias in value-added models," *American Economic Review*, 106, 393–99.

——— (2017): "Measuring the Impacts of Teachers: Reply to Rothstein," .

Clark, D., P. Martorell, and J. Rockoff (2009): "School Principals and School Performance. Working Paper 38." *National Center for Analysis of longitudinal data in Education research.*

Coelli, M. and D. A. Green (2012): "Leadership effects: School principals and student outcomes," *Economics of Education Review*, 31, 92–109.

Colonnelli, E., M. Prem, and E. Teso (2019): "Patronage and selection in public sector organizations," *CEPR Discussion Paper No. DP13697.*

Contreras, D., J. Rodriguez, and S. Urzua (2018): "The Return to Private Education: Evidence from School-to-Work Transitions," *Working Paper.*

Cowen, J. M. and M. A. Winters (2013): "Do charters retain teachers differently? Evidence from elementary schools in Florida," *Education Finance and Policy*, 8, 14–42.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009): "Dealing with limited overlap in estimation of average treatment effects," *Biometrika*, 96, 187–199.

Cullen, J. B., E. A. Hanushek, G. Phelan, and S. G. Rivkin (2016): "Performance Information and Personnel Decisions in the Public Sector: The Case of School Principals," Tech. rep., National Bureau of Economic Research.

Dal Bó, E., F. Finan, and M. A. Rossi (2013): "Strengthening state capabilities: The role of financial incentives in the call to public service," *The Quarterly Journal of Economics*, 128, 1169–1218.

Deserranno, E. (2019): "Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda," *American Economic Journal: Applied Economics*, 11, 277–317.

Dhuey, E. and J. Smith (2014): "How important are school principals in the production of student achievement?" *Canadian Journal of Economics/Revue canadienne d'économique*, 47, 634–663.

Di Liberto, A., F. Schivardi, and G. Sulis (2015): "Managerial practices and student performance," *Economic Policy*, 30, 683–728.

Duflo, E., P. Dupas, and M. Kremer (2011): "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya," *American Economic Review*, 101, 1739–74.

Eberts, R. W. and J. A. Stone (1988): "Student achievement in public schools: Do principals make a difference?" *Economics of Education Review*, 7, 291–299.

Elacqua, G. (2009): "For-profit schooling and the politics of education reform in Chile: When ideology trumps evidence," *Documento de Trabajo CPCE*, 5.

Fenizia, A. (2019): "Managers and Productivity in the Public Sector," *Working Paper*.

Ferraz, C. and F. Finan (2011): "Electoral accountability and corruption: Evidence from the audits of local governments," *American Economic Review*, 101, 1274–1311.

Finan, F., B. A. Olken, and R. Pande (2015): "The personnel economics of the state," Tech. rep., National Bureau of Economic Research.

Grissom, J. A., D. Kalogrides, and S. Loeb (2015): "Using student test scores to measure principal performance," *Educational Evaluation and Policy Analysis*, 37, 3–28.

Guerra, T. and N. Arcos (2012): "Ley SEP: Ranking de las 20 municipalidades más cuestionadas por Contraloría, available at: *https://ciperchile.cl/2012/05/28/ley-sep-ranking-de-la-20-municipalidades-mas-cuestionadas-por-contraloria/*," *CIPER Chile*.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin (2004): "Why public schools lose teachers," *Journal of human resources*, 39, 326–354.

Hanushek, E. A., S. G. Rivkin, and J. C. Schiman (2016): "Dynamic effects of teacher turnover on the quality of instruction," *Economics of Education Review*, 55, 132–148.

HARRIS, D. N. AND T. R. SASS (2006): "Value-added models and the measurement of teacher quality," *Unpublished manuscript.*

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor economics*, 24, 411–482.

HENDRICKS, M. D. (2014): "Does it pay to pay teachers more? Evidence from Texas," *Journal of Public Economics*, 109, 50–63.

HICKS, J. (1963): *The theory of wages*, Springer.

HOLLAND, P. W. (1986): "Statistics and causal inference," *Journal of the American statistical Association*, 81, 945–960.

HOLMSTROM, B. AND P. MILGROM (1987): "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica: Journal of the Econometric Society*, 303–328.

HORVÁTH, H. (2015): "Classroom Assignment Policies and Implications for Teacher Value-Added Estimation," *Working Paper.*

HSIEH, C.-T. AND M. URQUIOLA (2006): "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program," *Journal of public Economics*, 90, 1477–1503.

JACOB, B. A. AND L. LEFGREN (2007): "What do parents value in education? An empirical investigation of parents' revealed preferences for teachers," *The Quarterly Journal of Economics*, 122, 1603–1637.

——— (2008): "Can principals identify effective teachers? Evidence on subjective performance evaluation in education," *Journal of labor Economics*, 26, 101–136.

JOCHMANS, K. AND M. WEIDNER (2019): "Fixed-Effect Regressions on Network Data," *Econometrica*, 87, 1543–1560.

KALOGRIDES, D. AND S. LOEB (2013): "Different teachers, different peers: The magnitude of student sorting within schools," *Educational Researcher*, 42, 304–316.

KANE, T. J., D. F. MCCAFFREY, T. MILLER, AND D. O. STAIGER (2013): "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Research Paper. MET Project." *Bill & Melinda Gates Foundation.*

Kane, T. J. and D. O. Staiger (2008): "Estimating teacher impacts on student achievement: An experimental evaluation," Tech. rep., National Bureau of Economic Research.

Kremer, M. and E. Maskin (1996): "Wage inequality and segregation by skill," Tech. rep., National bureau of economic research.

Lavy, V. and A. Boiko (2017): "Management Quality in Public Education: Superintendent Value-Added, Student Outcomes and Mechanisms," Tech. rep., National Bureau of Economic Research.

Lazear, E. P. and S. Rosen (1981): "Rank-order tournaments as optimum labor contracts," *Journal of political Economy*, 89, 841–864.

Lazear, E. P. and K. L. Shaw (2007): "Personnel economics: The economist's view of human resources," *Journal of economic perspectives*, 21, 91–114.

Logan, J. A. (1996): "Opportunity and choice in socially structured labor markets," *American Journal of Sociology*, 102, 114–160.

Loyalka, P., S. Sylvia, C. Liu, J. Chu, and Y. Shi (2019): "Pay by design: Teacher performance pay design and the distribution of student achievement," *Journal of Labor Economics*, 37, 621–662.

Manning, A. (2011): "Imperfect Competition in the Labor Market," in *Handbook of labor economics*, Elsevier, vol. 4, 973–1041.

Marshall, A. (1890): *Principles of economics*, Macmillan.

McCormack, J., C. Propper, and S. Smith (2014): "Herding cats? Management and university performance," *The Economic Journal*, 124, F534–F564.

Mundlak, Y. (1978): "On the pooling of time series and cross section data," *Econometrica: journal of the Econometric Society*, 69–85.

Neilson, C. (2019): "Targeted Vouchers, Competition among Schools and the Academic Achievement of Poor Students," *Working Paper*.

Paulo, S., F. Ariel, G. J. Sandra, and R. Thomas (2017): *OECD Reviews of School Resources: Chile 2017. Chapter 4*, OECD Publishing.

Peteck, N. and N. Pope (2019): "The Multidimensional Impact of Teachers on Students," *Working Paper*.

Ransom, M. R. and D. P. Sims (2010): "Estimating the firm's labor supply curve in a "new monopsony" framework: Schoolteachers in Missouri," *Journal of Labor Economics*, 28, 331–355.

Rasul, I. and D. Rogger (2018): "Management of bureaucrats and public service delivery: Evidence from the nigerian civil service," *The Economic Journal*, 128, 413–446.

Robinson, J. (1969): *The economics of imperfect competition*, Macmillan.

Romano, J. P. and M. Wolf (2005): "Stepwise multiple testing as formalized data snooping," *Econometrica*, 73, 1237–1282.

Ronfeldt, M., S. Loeb, and J. Wyckoff (2013): "How teacher turnover harms student achievement," *American Educational Research Journal*, 50, 4–36.

Roth, A. E. and M. Sotomayor (1990): "a Study in Game-theoretic Modeling and Analysis," *Econometric Society Monographs*, 18.

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

——— (2015): "Teacher quality policy when supply matters," *American Economic Review*, 105, 100–130.

Roy, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford economic papers*, 3, 135–146.

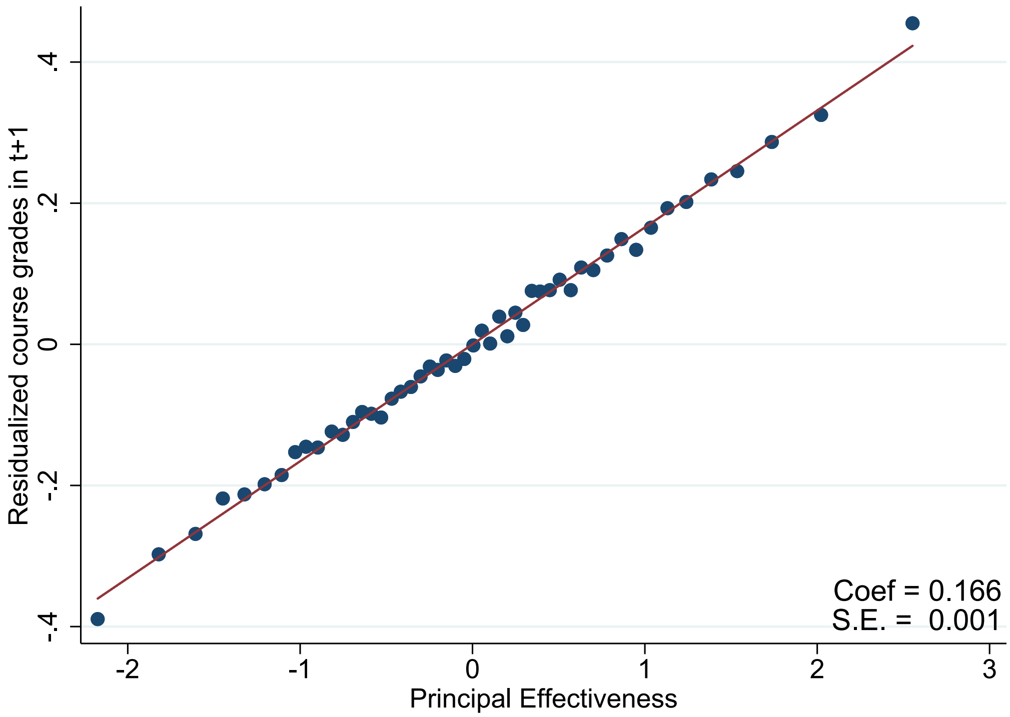
Staiger, D. O. and J. E. Rockoff (2010): "Searching for effective teachers with imperfect information," *Journal of Economic perspectives*, 24, 97–118.

Valverde, G. A. (2004): "Curriculum convergence in Chile: The global and local context of reforms in curriculum policy," *Comparative Education Review*, 48, 174–201.

Wooldridge, J. M. (2010): "Correlated random effects models with unbalanced panels," *Michigan State University, Department of Economics.*

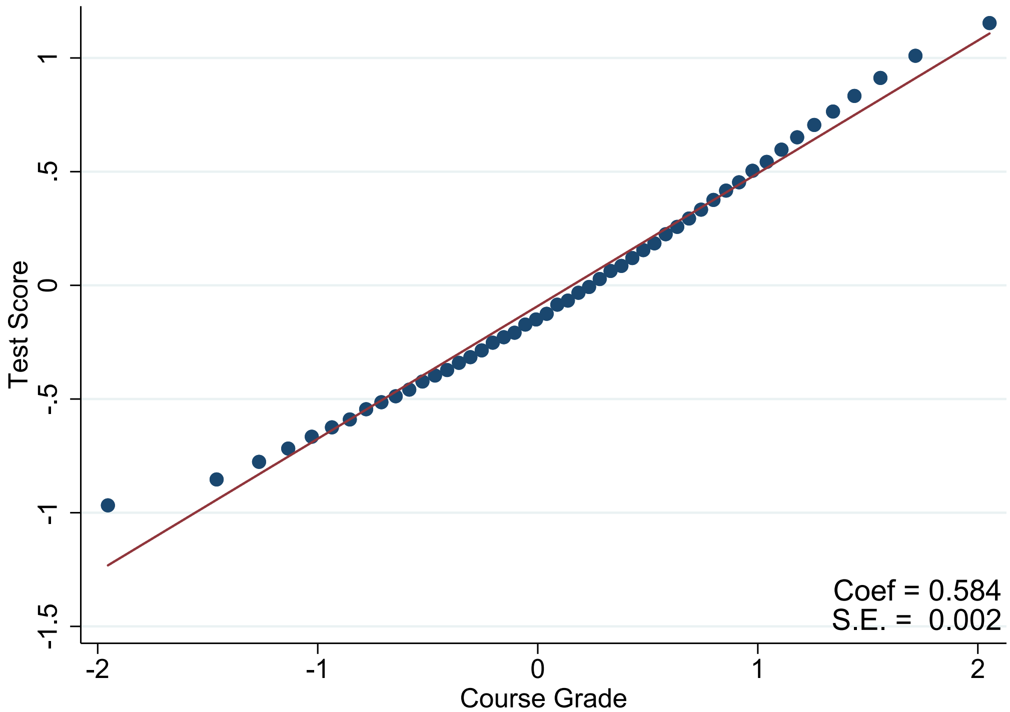
Figures and Tables

Figure 1: Residual course grades and principal fixed effects



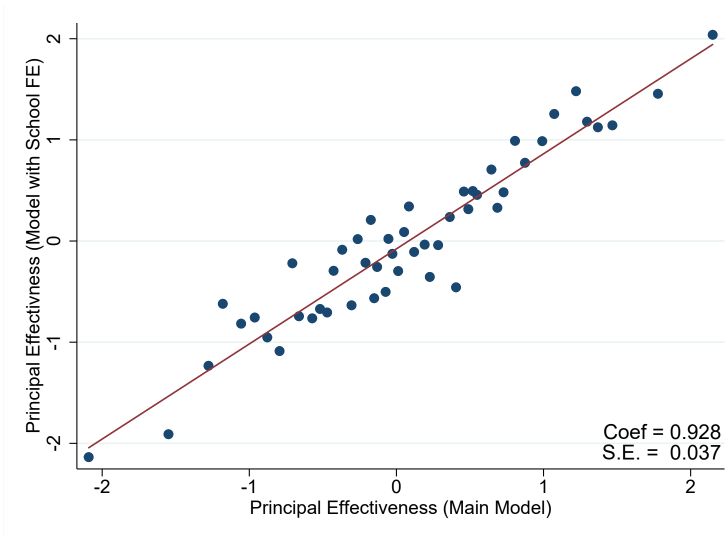
Notes: We residualized course grades with respect to the baseline vector of controls used in our main specification, but excluding principal fixed effects. Then we run residualized course grades on our estimated measure of principal effectiveness. The regression includes grade and subject fixed effects. The estimation sample is the same used to obtain principal fixed effects and it includes 8,687,086 observations. Bootstrap standard errors clustered at the school-grade level (100 replications).

Figure 2: Relationship between course grades and Test Score

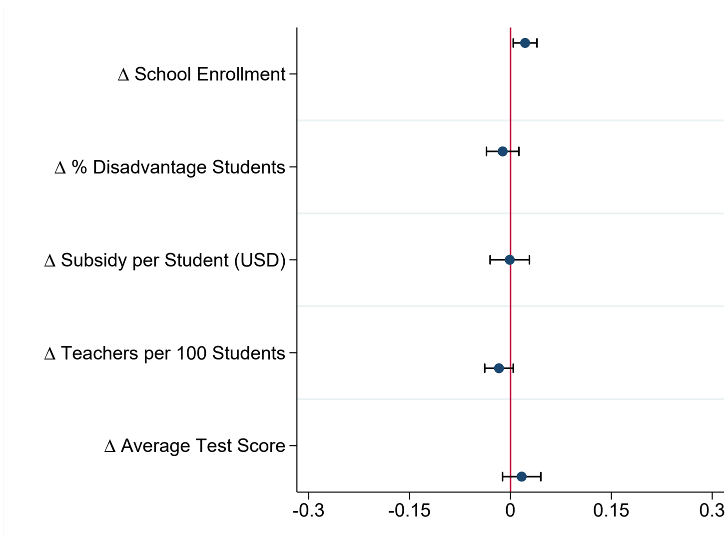


Notes: This figure shows a regression of test scores on course grades, including subject-grade and school fixed effects. The estimation sample includes 1,061,231 students for whom we observe test score and course grades contemporaneously for Math and Spanish. Robust standard errors are clustered at the school level.

Figure 3: Schools as confounder



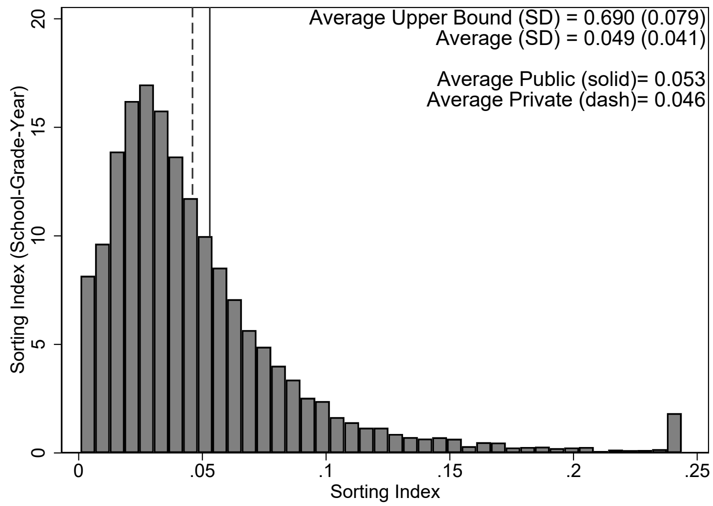
A. Comparison of alternative measures of principal effectiveness



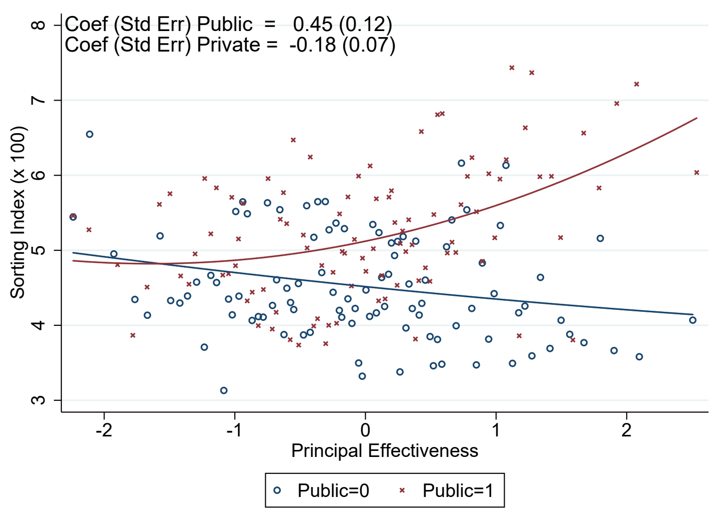
B. Δ Principal effectiveness and Δ School characteristics

Notes: We focus on 300 school principals who switched schools at some point and that we observe for at least two periods in each school. Using this sample, we estimate our main specification but including school fixed effects instead of correlated random effects. Panel A shows the relationship between the measure of principal effectiveness obtained from this model and our main measure of principal effectiveness. Panel B also considers principals switching schools and shows how changes in the measure of principal effectiveness relates to changes in predetermined school characteristics (before 2010), as described by equation (2).

Figure 4: School-level sorting



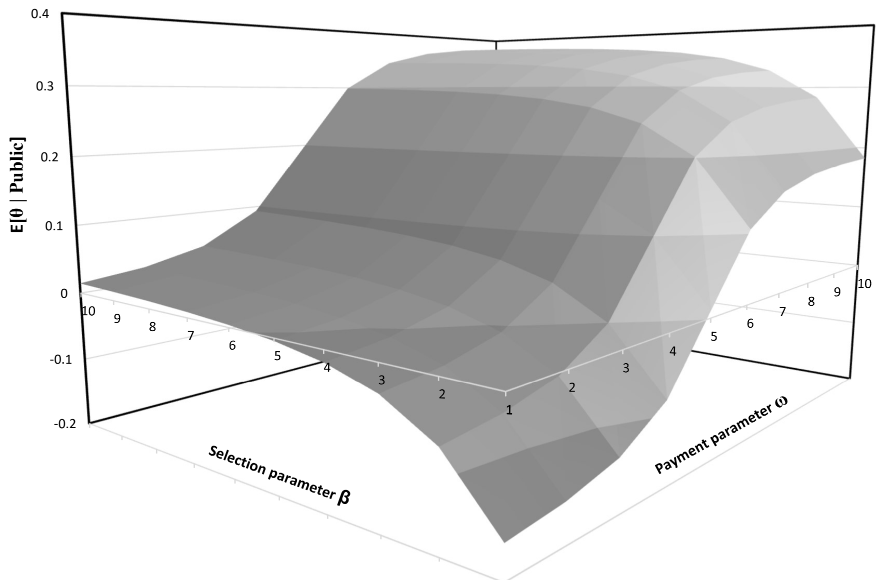
A. Sorting Index



B. Sorting Index and Principal Quality

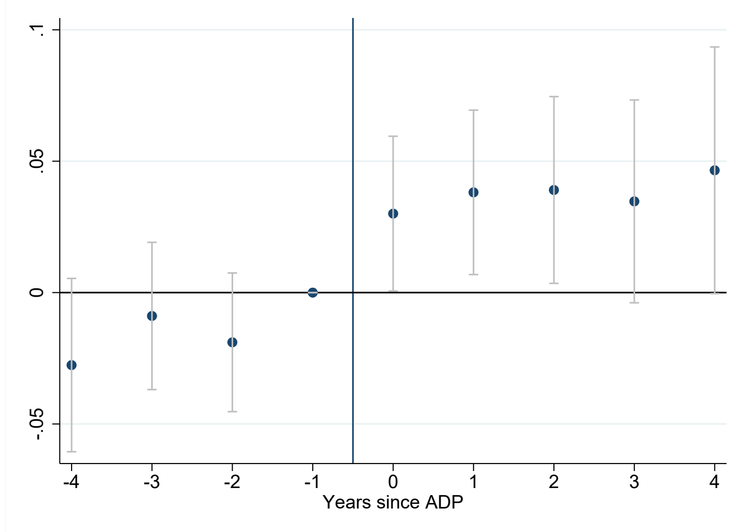
Notes: Panel A plots the histogram of a school-year measure of sorting of students across classrooms. Specifically, it plots the τ_{st} coefficients from several Kremer and Maskin (1996) type of regressions described in section 4. The top right of the figure presents the actual between-classrooms (and within school-grade) sorting and a theoretical upper (UB) of sorting. To construct this UB, we rank students within a school-grade, according to their course grades, and then we sort them across classrooms in the order determined by their ranking. Panel B shows the relationship between our measure of sorting and standardized principal effectiveness, by type of school. Bootstrapped standard errors (100 replications) are clustered at the principal level.

Figure 5: Model simulation

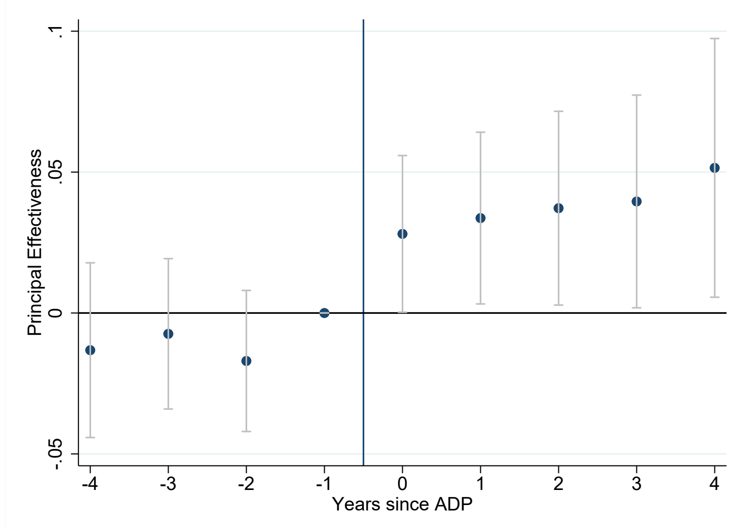


Notes: This plot shows how the allocation of principal effectiveness depends on the selection and payment parameters of the theoretical model presented in section 6. To construct this figure, we created a grid for β_p and w_p from 1 to 10, and compute equation (15) for each cell of this grid. We set all other parameters $(h_p, h_v, \beta_{0p}, \beta_{0v}, w_v, \beta_v)$ at constant baseline levels (1,1,1,1,5,5).

Figure 6: Event Study



A. Without Controls



B. With Controls

Notes: This figure shows the impact of the new selection system on the effectiveness of public schools' principals. Specifically, it plots the point estimates and 95% confidence intervals estimated from equation (17). Panel A only considers school and year fixed effects. Panel B considers school and year fixed effects and also controls by school and municipality characteristics during the pre-reform period (measured in 2010), interacted with year dummies.

Table 1: Summary statistics

	(1) Mean	(2) Standard deviation	(3) Median	(4) 10th pctile	(5) 90th pctile
Panel A: Student characteristics					
Math course grade	5.2	0.9	5.1	4.0	6.4
Spanish course grade	5.2	0.8	5.2	4.2	6.3
Math test scores	255.6	49.7	256.9	189.6	319.0
Spanish test scores	253.5	50.3	255.9	184.1	317.9
Grade retention (%)	4.6	20.9	0.0	0.0	0.0
Panel B: School characteristics					
Enrollment	299.3	397.4	153.0	3.0	817.0
Class size	20.0	12.9	21.0	2.0	37.0
Annual subsidy per student (USD)	2424.3	3704.6	1446.5	840.8	4089.8
Share of disadvantaged students	49.0	35.1	58.0	0.0	93.0
Teachers per hundred students	8.0	17.1	5.7	2.6	14.3
Rural school	46.6	49.9	0.0	0.0	100.0
School attendance	86.1	9.5	87.8	76.2	95.1
Panel C: Principal characteristics					
Wage (USD)	2738.8	1982.4	2525.3	1587.0	3920.8
% Base salary	42.5	18.9	35.9	24.0	72.3
% Bonus	16.0	18.9	7.5	0.8	48.5
% Statutory	41.5	22.7	43.3	10.5	69.0
Tenure	27.3	12.2	29.0	10.0	42.0
Permanent contract	90.6	29.2	100.0	100.0	100.0
Age	54.2	9.9	55.0	40.0	65.0
Female	52.5	49.9	100.0	0.0	100.0

Notes: This table presents summary statistics for students, schools, and principals. We only include subsidized private and public schools.

Table 2: Differences between public and private schools

	(1) Public	(2) Private	(3) Difference
Panel A: Student characteristics			
Math course grade (standardized)	-0.03	0.02	-0.05
	(0.99)	(1.01)	(0.00)***
Spanish course grade (standardized)	-0.05	0.03	-0.09
	(1.02)	(0.98)	(0.00)***
Math test scores (standardized)	-0.23	0.16	-0.38
	(1.00)	(0.97)	(0.00)***
Spanish test scores (standardized)	-0.18	0.11	-0.30
	(1.00)	(0.98)	(0.00)***
Grade retention (%)	4.92	4.33	0.59
	(21.62)	(20.35)	(0.00)***
Panel B: School characteristics			
Enrollment	215.90	415.69	-199.79
	(303.61)	(475.76)	(0.00)***
Class size	17.17	22.47	-5.30
	(11.76)	(13.39)	(0.00)***
Annual subsidy per student (USD)	2977.05	1628.20	1348.85
	(4302.89)	(2398.08)	(0.00)***
Share of disadvantaged students	57.30	36.72	20.58
	(32.88)	(34.72)	(0.00)***
Teachers per hundred students	9.15	6.38	2.76
	(12.44)	(22.14)	(0.00)***
Rural school	62.40	24.45	37.95
	(48.44)	(42.98)	(0.00)***
School attendance	85.78	86.51	-0.73
	(9.23)	(9.90)	(0.00)***
Panel C: Principal characteristics			
Wage (USD)	2593.73	2896.42	-302.69
	(2301.79)	(1548.07)	(0.00)***
% Base salary	35.34	50.25	-14.92
	(13.48)	(20.74)	(0.00)***
% Bonus	10.26	22.23	-11.97
	(15.11)	(20.64)	(0.00)***
% Statutory	54.40	27.51	26.88
	(18.37)	(18.24)	(0.00)***
Tenure	28.70	25.69	3.01
	(11.72)	(12.47)	(0.00)***
Permanent contract	87.79	93.63	-5.85
	(32.75)	(24.42)	(0.00)***
Age	54.85	53.58	1.27
	(8.84)	(10.97)	(0.00)***
Female	44.73	60.85	-16.12
	(49.73)	(48.82)	(0.00)***

Notes: This table shows the differences between private and public schools in terms of student, school, and principal characteristics. Columns 1 and 2 present the average and standard deviation, respectively; and column 3 presents the difference between both and the p-value of the difference (in parenthesis). We consider only subsidized private and public schools.

Table 3: Assignment of students to principals

Panel A: Regression of Gain Achievement

	(1)	(2)	(3)	(4)
	Course grade at school s_0			
	Math	Spanish	Math	Spanish
$\hat{\theta}_{p(i)}$ at school s_0	0.112***	0.110***	0.120***	0.116***
	(0.009)	(0.008)	(0.009)	(0.009)
$\hat{\theta}_{p(-i)}$ at school s_1			-0.016**	-0.013*
			(0.007)	(0.007)
Observations	116,349	119,778	116,349	119,778

Panel B: Regression with Lagged Achievement Control

	(1)	(2)	(3)	(4)
	Course grade at school s_0			
	Math	Spanish	Math	Spanish
$\hat{\theta}_{p(i)}$ at school s_0	0.182***	0.182***	0.177***	0.175***
	(0.008)	(0.007)	(0.008)	(0.008)
$\hat{\theta}_{p(-i)}$ at school s_1			0.011	0.015**
			(0.007)	(0.006)
Observations	116,349	119,778	116,349	119,778

Notes: This table shows the results from the validation exercise discussed in section 3. All specifications include grade fixed effects. Bootstrapped standard errors (100 replications) clustered by school of origin.

Table 4: Teacher turnover and principal effectiveness

Dependent variable: 1 if teacher leaves the school next period	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Teachers by age group			Teachers by Value-Added		
		$\in [22, 37)$	$\in [37, 52)$	$\in [52, 67)$	Low VA	Med. VA	High VA
All Schools							
Principal Effectiveness	-0.003***	-0.004***	-0.002**	-0.001	-0.000	-0.003**	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log Teachers' Pay	-0.061***	-0.061***	-0.066***	-0.050***	-0.064***	-0.050***	-0.067***
	(0.006)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)
Mean Exit rate	0.108	0.150	0.0941	0.0486	0.0936	0.127	0.100
Public Schools							
Principal Effectiveness	-0.001	-0.001	-0.001	-0.001	0.000	0.001	-0.003*
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Log Teachers' Pay	-0.031***	0.020	-0.058***	-0.042***	-0.048***	-0.000	-0.029
	(0.011)	(0.019)	(0.015)	(0.013)	(0.016)	(0.020)	(0.019)
Mean Exit rate	0.0961	0.141	0.0965	0.0475	0.0834	0.117	0.0885
Private Schools							
Principal Effectiveness	-0.003**	-0.005***	-0.002*	0.001	0.000	-0.004*	-0.007***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Log Teachers' Pay	-0.077***	-0.082***	-0.075***	-0.053***	-0.076***	-0.072***	-0.085***
	(0.007)	(0.012)	(0.009)	(0.012)	(0.010)	(0.012)	(0.013)
Mean Exit rate	0.118	0.155	0.0919	0.0507	0.103	0.134	0.111

Notes: This table presents the results from the linear probability model described by equation (4). For the estimation we focus on prime age teachers, between 22 and 62 years old, from 2008-2016. The dependent variable equals 1 if a teacher experienced a job-to-job or job-to-non-teaching employment transition. School-level controls include the average wage paid to teachers, the share of disadvantaged students, school size, and type of school (public and rural indicators). Teacher-level controls include gender, age, age squared, and type of contract. All specifications include year and municipality fixed effects. Bootstrap standard errors (100 replications) clustered at the principal level are in parenthesis.

Table 5: Principal effectiveness and perceptions of the school community

	(1)	(2)	(3)	(4)	(5)	(6)
	β Coef.	Std. Error	R-W p-val	Perm. p-val	$\bar{Y}$	Obs.
Panel A: Teachers' Responses						
Does a good job	0.022***	(0.005)	0.000	0.000	0.470	6288
Promotes good climate	0.020***	(0.005)	0.000	0.000	0.551	6192
Is a good manager	0.018***	(0.004)	0.000	0.000	0.503	7412
Can be trusted	0.012***	(0.004)	0.020	0.010	0.539	6276
Makes good decisions	0.012***	(0.004)	0.010	0.000	0.484	7413
Is effective	0.009**	(0.004)	0.020	0.040	0.476	7407
Is good at communicating	0.019***	(0.005)	0.000	0.000	0.529	6293
Engages teachers	0.013***	(0.005)	0.000	0.000	0.466	7386
Engages parents	0.017***	(0.004)	0.000	0.000	0.488	7413
Knows teacher needs	0.012***	(0.004)	0.010	0.010	0.462	7406
Knows student needs	0.019***	(0.005)	0.000	0.000	0.527	6291
Includes teachers	0.010***	(0.004)	0.010	0.015	0.503	8281
Panel B: School Complaints						
Bullying/Discrimination	-0.101***	(0.026)	0.000	0.020	0.612	4831
Denied enrollment	-0.029***	(0.010)	0.010	0.065	0.163	4831
Poor infrastructure	-0.040*	(0.022)	0.010	0.045	0.112	4831
Teacher absenteeism	-0.017**	(0.008)	0.178	0.190	0.092	4831
School accidents	-0.006	(0.010)	0.782	0.350	0.082	4831
Charges extra fees	0.004	(0.005)	0.782	0.095	0.035	4831
Resource accountability	-0.007	(0.011)	0.782	0.270	0.066	4831

Notes: In Panel A, we first create a dummy at the survey respondent level where the variable takes a value of one if the survey response is "highly agree" with the statement. In cases when the survey had 5 or 4 options we always use the highest number to create the dummy. Then, we take the average across respondents at the school-year level and assign this to a principal. In Panel B, we consider all complaints (per 100 students) filed against a school (as recorded by the *Superintendencia de Educación*) and we sum them across years to then assign them to a principal. All models include municipality fixed effects. Columns 1 and 2 report the estimated coefficients and bootstrapped standard errors (100 replications) from specification (6). Column 3 presents p-values adjusted for multiple testing using the step-down procedure of Romano and Wolf (2005). Column 4 reports the results from a permutation test for which we randomly reshuffled principal fixed effects 1,000 times. The p-value of the test is calculated as the proportion of sampled permutations where the value of $\hat{\beta}_s$ was greater than or equal to our estimate $\hat{\beta}$. Finally, columns 5 and 6 present the mean of the dependent variable and the number of observations used for estimation.

Table 6: Principal compensation and principal effectiveness

Panel A: All Wages

	Private Schools		*Public Schools*	
	ln(Wage)	ln(Wage)	ln(Wage)	ln(Wage)
Principal Effectiveness	0.038***	0.038***	0.009**	0.007**
	(0.012)	(0.012)	(0.004)	(0.003)
Tenure		0.009***		0.011***
		(0.003)		(0.002)
$Tenure^2$		-0.000**		-0.000**
		(0.000)		(0.000)
Female		-0.116***		-0.015**
		(0.018)		(0.006)
Observations	5,917	5,917	6,314	6,314
R-squared	0.212	0.274	0.415	0.537

Panel B: Wage Components

	Private Schools		*Public Schools*	
	ln(Base)	$\ln\left(\frac{\text{Wage}}{\text{Base}}\right)$	ln(Base)	$\ln\left(\frac{\text{Wage}}{\text{Base}}\right)$
Principal Effectiveness	0.022*	0.016**	0.002	0.004**
	(0.011)	(0.007)	(0.003)	(0.002)
Tenure	0.003	0.006***	0.010***	0.000
	(0.003)	(0.002)	(0.002)	(0.001)
$Tenure^2$	0.000	-0.000***	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.065***	-0.050***	-0.013*	-0.002
	(0.019)	(0.017)	(0.007)	(0.003)
Observations	5,917	5,917	6,314	6,314
R-squared	0.184	0.150	0.525	0.517

Notes: This table presents the results from specification (7). We focus on a sample of principals for whom we have an standardized measure of effectiveness and detailed wage data from 2015 to 2017. All specifications include year and municipality fixed effects. Bootstrapped standard errors (100 replications) clustered at the principal level are in parenthesis.

Table 7: Effect of ADP selection on principal effectiveness

Dependent variable: Standardized school principal effectiveness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					Truncated sample	Within ADP	Placebo I	Placebo II
ADP turnover	0.043***	0.043***	0.044***	0.047***	0.048***	0.042***		
	(0.012)	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)		
	[0.021]	[0.021]	[0.021]	[0.021]	[0.022]	[0.016]		
Principal turnover							-0.009	0.010
							(0.059)	(0.011)
							[0.073]	[0.020]
Observations	21,668	21,668	21,668	21,668	19,403	12,760	8,960	
R-squared	0.849	0.849	0.849	0.843	0.843	0.779	0.971	0.890
# of Schools	3167	3167	3167	3167	2818	1663	2549	3644
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	No	Yes	No	No	No	No	No
Municipality controls	Yes	Yes	No	No	No	No	No	No

Notes: This table presents the effects of the new selection system (ADP) on the standarized measure of principal effectiveness discussed in section 3. "ADP turnover" is a dummy that takes the value one after the first time a school selects a principal under the ADP system. Columns 1 to 4 estimate the regressions described by equation (16). Column 5 follows Crump et al. (2009) and truncates the sample based on a propensity score that estimates the probability that a school selects a principal under the ADP system. The optimal cut-off in our case is 11%. Column 6 estimates the main regression only within schools that selected a principal under the ADP system. Column 7 shows a placebo exercise where "Principal turnover" is a dummy that takes the value one after a principal turnover in a public school in the period 2009-2010 (pre-ADP reform). The number of schools who had a principal turnover in 2009 or 2010 is 872. Column 8 shows a similar placebo exercise where we focus on school turnover after 2012 but in private schools. The number of private schools that had a turnover after 2012 is 2,223. Robust standard errors in parenthesis, and robust standard errors clustered at school level in brackets.

Table 8: Comparison Table

Dependent Variable: Principal Effectiveness (standardized)

	$\hat{\theta}_p$ **Course Grades**			$\hat{\theta}_p$ **Test Scores**		
	β Coef	Std. Error	Obs.	β Coef	Std. Error	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Resid. Achievement (Standardized)						
Course Grades / Test Scores	0.166***	(0.000)	8,687,086	0.239***	(0.005)	235,434
Teachers' Exit Rate						
All Teachers	-0.003***	(0.001)	351,633	-0.003***	(0.001)	277,721
Young Teachers	-0.005***	(0.002)	148,284	-0.004***	(0.002)	120,605
High VA Teachers†	-0.006***	(0.002)	108,826	-0.001	(0.001)	85,639
Log Wages						
All Schools	0.013***	(0.005)	12,231	0.032***	(0.006)	9,688
Public Schools	0.007**	(0.003)	6,314	0.008**	(0.004)	4,664
Private Schools	0.038***	(0.010)	5,917	0.036***	(0.011)	5,024
ADP Selection (Dif-in-Dif)						
With Controls	0.043***	(0.012)	21,668	0.043***	(0.014)	12,904
Without Controls	0.046***	(0.011)	22,405	0.029**	(0.014)	13,235

Notes: This table compares our main results with those obtained when we use an alternative measure of principal effectiveness based on test scores. †: Notice that we cannot estimate teacher value-added using test scores as students do not take exams every year. Thus, we use our main measure of teacher value-added based on course grades.

Appendix A: Teacher Value-Added

Following Kane and Staiger (2008), we construct a measure of teacher value-added based on the residual of the model described by equation (1). Specifically, we assume that the error term can be decomposed into four components, as follows:

$$e_{it+1} = \mu_j + \rho_c + \psi_s + \varepsilon_{it+1},$$

where μ_j stands for teacher value-added, ρ_c and ψ_s stand for exogenous class and school level shocks, and ε_{it+1} represents idiosyncratic student level variation. Let $\varepsilon_{it} = \rho_c + \psi_s + \varepsilon_{it+1}$ denote the unobserved error in course grade unrealed to teacher quality. To obtain unbiased estimates of μ_j, we proceed in three steps:

- First, we estimate the variance of the teacher, classroom, school, and student components of the residual in equation (1). To obtain an estimate of the variance of the student component we use the within-classroom variance in e_{it+1}:

  $$\hat{\sigma}_e^2 = \text{Var}(e_{i(j)t+1} - \bar{e}_{jt+1}).$$

  Similarly, we estimate the variance of the school component using the within-school variance in e_{it+1}:

  $$\hat{\sigma}_\psi^2 = \text{Var}(e_{i(s)t+1} - \bar{e}_{st+1}).$$

  Like in Chetty et al. (2014), we scale $\hat{\sigma}_e^2$ and $\hat{\sigma}_\psi^2$ to correct the degrees of freedom for the fact that we have already estimated parameters to form the residual. Then, we estimate the variance in the teacher component with the covariance between the average residual in a teacher's class in year $t + 1$ and year t:[32]

  $$\hat{\sigma}_\mu^2 = \text{Cov}(\bar{e}_{jt+1}, \bar{e}_{jt}).$$

  Finally, we estimate the variance of the classroom component as the remainder:

  $$\hat{\sigma}_\rho^2 = \text{Var}(e_{ijt+1}) - \hat{\sigma}_\mu^2 - \hat{\sigma}_\psi^2 - \hat{\sigma}_e^2.$$

[32] To deal with teachers that teach more than one class per year, we collapse the data to the teacher-year level and construct precision-weighted averages of classroom-average residuals within a teacher-year (with weights given by the number of students in each classroom).

- Second, as in Kane and Staiger (2008), we form a weighted average of the average classroom residuals for each teacher ($\bar{e}_{jt+1}$) that is a minimum variance unbiased estimate of μ_j for each teacher (so that weighted average has maximum reliability). Data from each classroom is weighted by its precision (the inverse of the variance), with larger classrooms having less variance and receiving more weight:

$$\bar{e}_j = \sum_t w_{jt}\bar{e}_{jt+1},$$

where $w_{jt} = h_{jt}/\sum_t h_{jt}$, and

$$h_{jt} = \frac{1}{\text{Var}(\bar{e}_{jt+1} \mid \mu_j)} = \frac{1}{\hat{\sigma}^2_\rho + (\frac{\hat{\sigma}^2_e}{n_{jt}})}.$$

- Finally, we construct an empirical Bayes estimator of each teacher's value added $\bar{e}_j$ by multiplying the weighted average of classroom residuals by an estimate of its reliability:

$$\text{VA}_j = \bar{e}_j \times \left(\frac{\hat{\sigma}^2_\mu}{\hat{\sigma}^2_\rho + (\sum_t h_{jt})^{-1}}\right)$$

The quantity in parenthesis represents the shrinkage factor, and reflects the reliability of $\bar{e}_j$ as an estimate of μ_j, where the reliability is the ratio of signal variance to total variance. Note that the total variance is the sum of signal variance and estimation error variance, and the estimation variance for $\bar{e}_j$ can be shown to be $(\sum_t h_{jt})^{-1}$.

Appendix B: Principal Compensation in Public Schools

Do higher wages attract more effective principals into public schools? Although this question is of high practical relevance, empirical progress has met with at least two important hurdles. First, it is difficult to measure an individual's quality, and second, different wages for a given position are not typically assigned exogenously. In this section, we attempt to answer that question by leveraging quasi-experimental variation in principal compensation together with our measure of principal effectiveness.

Related research on this topic has shown that higher wages can attract higher quality candidates into the workforce of the state (see Dal Bó et al., 2013), as well as better politicians into the legislative body (Ferraz and Finan, 2011), pointing to a positive relationship between compensation and the development of state capabilities. In this spirit, the *Quality and Equity in Education* Law passed in Chile in 2011 increased the compensation for school principals using an arbitrary rule that depends on the school size, e.g., total enrollment, and the concentration of disadvantaged students, e.g., share of students eligible for a voucher targeted to the poor. Since 2012, the compensation of the principal increases discontinuously in schools with more than 60% of disadvantaged students, and around the school-size thresholds of 400, 800 or 1,200 enrolled students.

In what follows, we focus our empirical analysis on the cutoff with the largest mass around it: the school-size threshold of 400 enrolled students. Our analysis uses administrative data on wages between 2015 and 2017 and digitized data on raw wages from 2009 to 2014. The former administrative data was obtained directly from the Superintendency of Education and the latter data was obtained from Chilean municipalities, via a Freedom-of-Information request, and it is only available for a subset of schools.

Empirical strategy and results

We are interested in the following relationship between wages and principal effectiveness:

$$\hat{\theta}_p = \beta \ln(wage_p) + f(r_p) + e_p, \tag{18}$$

where $\hat{\theta}_p$ is the standardized version of the measure of principal effectiveness discussed in section 3, $\ln(wage_p)$ is the logarithm of the average monthly wage paid to the school principal, $f(r_p)$ is a function of the school size, and e_p is an error term. A potential threat to identification in this setting is that the wage paid by a given school to the principal is not exogenous. If that is the case, ordinary least squares (OLS) estimates of equation (18) will not recover the causal effect of wages. We attempt to address this concern by estimating a two-stage least squares model (2SLS) with the following first-stage equation:

$$\ln(wage_p) = \pi Z_p + g(r_p) + v_p, \tag{19}$$

where $Z_p = 1(r_p > 0)$ is a dummy that equals 1 if the school has at least 400 students, and $g(r_p)$ is a linear function of the school size $r_p = (\text{School Size} - 400)$, at both sides of the threshold.

We start by noticing that covariates are balanced between "treated" and "untreated" schools. Schools right above the cutoff look much like schools below. Table B.1 presents a simple comparison of baseline characteristics for schools at both sides of the cutoff. These estimated coefficients come from separate regressions of each baseline characteristic on the initial eligibility indicator Z_p, substituting $\ln(wage_p)$ by the corresponding characteristic in equation (19). Column 3 of this table presents the p-values from the permutation test for the difference proposed by Canay and Kamat (2017). We do not find any statistically significant difference. However, visual inspection of the density of the running variable in Figure B.1 seems to indicate some manipulation of total enrollment.[33]

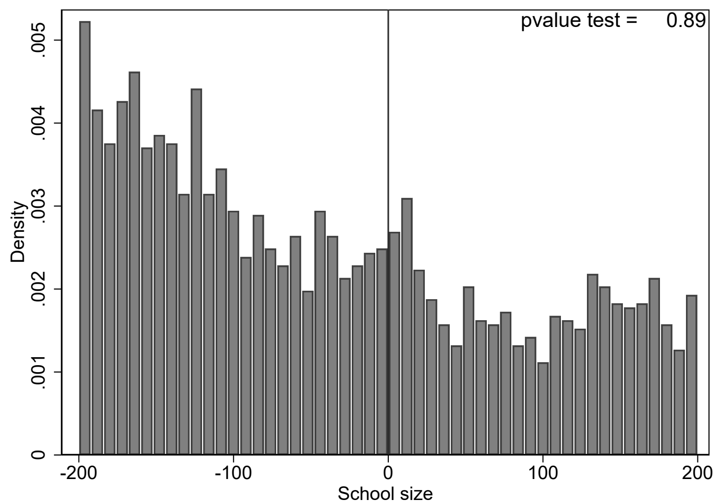
Our main results are presented in Table B.2. For estimation, we consider a sample of schools within an optimal bandwidth of 96 students around the cutoff and a linear school size polynomial.[34] Columns 1 to 3 focus on the post-reform period (2012-2017). Column 1 shows the first stage. We find that crossing the school-size threshold increases wages by 6.2%, a statistically significant effect but with an F-test slighlty below 10. Column 2 suggests that schools that are marginally above the school-size threshold have principals that are 0.16 standard deviations more effective, compared to schools marginally below the threshold. The results from our 2SLS approach are presented in column 3. We estimate a large but imprecise effect. A 10% increase in wages is associated to an average increase in principal effectiveness of 0.26 standard deviations, but at a 5% level of risk, this effect ranges from

[33]We passed the test proposed by Cattaneo et al. (2018), which fails to reject the null hypothesis of equal densities around the cutoff.

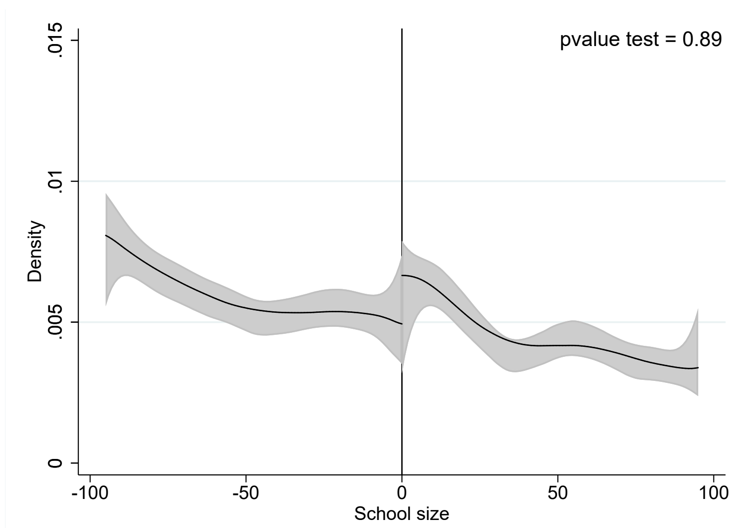
[34]We select the optimal bandwidth based on the reduced form regression, following Calonico et al. (2014). Table B.3 shows the robustness of our results to different bandwidths. However, as shown by Table B.4, the statistical significance of our results is sensible to the degree of the polynomial.

-0.05 to 0.57. Finally, we present a placebo exercise in which we estimate the first stage and reduced form models but during the pre-reform period (2009-2011). Columns 4 and 5 show these results. Reassuringly, we find that before the Law of *Quality and Equity in Education* of 2011, crossing the school-size threshold had a zero effect both on wages and principal effectiveness.

Figure B.1: Manipulation of school enrollment



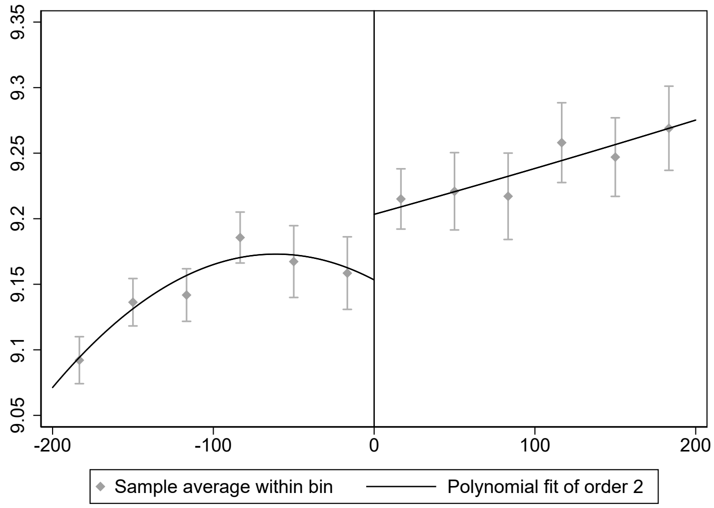
A. Manipulation: Histogram



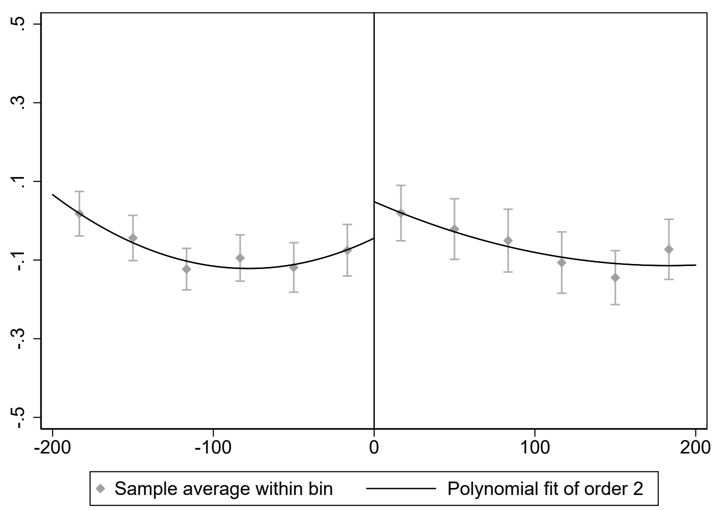
B. Manipulation: Density

Notes: Panels A and B present the histogram and density of the running variable: school enrollment. The vertical red line shows the cut-off at 400 students. At the top right of the figure, we present the p-value for the manipulation test around the cut-off proposed by Cattaneo et al. (2018).

Figure B.2: Wages and principal effectiveness



A. Effect on Wages



B. Effect on Principal Effectiveness

Notes: Panels A and B present the histogram and density of the running variable: school enrollment. The vertical red line shows the cut-off at 400 students. At the top right of the figure, we present the p-value for the manipulation test around the cut-off proposed by Cattaneo et al. (2018). Panels B and C present mean wages and principal effectiveness around the cut-off. Each dot represents the average within the bin, while the vertical lines shows the confidence interval at 95%.

Table B.1: Balance in observables around the cut-off

	(1)	(2)	(3)
	Unrestricted sample	Restricted sample	
	OLS	OLS	Canay & Kamat p-value
Panel A: School characteristics			
Rural School	0.13***	0.02	0.70
	(0.01)	(0.03)	
Share of poor students	0.99	-1.79	0.41
	(1.15)	(2.63)	
4rd grade Spanish test score	0.08**	-0.01	0.50
	(0.03)	(0.07)	
4rd grade maths test score	-0.01	-0.00	0.52
	(0.03)	(0.06)	
H-S grad Spanish test score	-0.05*	0.07	0.93
	(0.03)	(0.06)	
H-S grad math test score	-0.05*	0.09	0.60
	(0.03)	(0.06)	
Panel B: Municipality characteristics			
Share of HH in extreme monetary poverty	0.31**	-0.28	0.65
	(0.15)	(0.31)	
Average HH per capita income	0.03	-0.03	0.71
	(0.04)	(0.07)	
Unemployment rate	0.03	-0.34	0.11
	(0.19)	(0.48)	
Average years of schooling	-0.13**	0.09	0.43
	(0.05)	(0.11)	

Notes: This table shows the regression presented in equation (19) for different school and municipality characteristics measured before the reform. Column 1 presents the results for the full sample, while column 2 shows the result using the optimal bandwidth proposed by Calonico et al. (2014). Column 3 presents the p-value suggested by Canay and Kamat (2017) for the test of the difference around the cut-off. Robust standard errors are presented in parenthesis.

Table B.2: Wages and principal effectiveness

Dependent variable: Standardized school principal effectiveness

	(1)	(2)	(3)	(4)	(5)
	Post-Reform			Pre-Reform	
	Wage	Principal effectiveness	Principal effectiveness	Wage	Principal effectiveness
Log wage			2.597*		
			(1.572)		
Above threshold	0.062***	0.162**		-0.006	-0.034
	(0.026)	(0.099)		(0.036)	(0.112)
Observations	1,759	1,759	1,759	478	478
F excl instr			9.038		

Notes: This table presents the results from equation (19). Log wage is the logarithm of of monthly wages, and principal effectiveness is a standardize measure of our estimated principal fixed effect discussed in section 3. We select the optimal bandwidth from the reduced form regression following Calonico et al. (2014). This optimal bandwidth is 96. All regressions include a linear polynomial around the cut-off. Robust standard errors are presented in parenthesis.

Table B.3: Wages and principal effectiveness: Robustness to bandwidth selection

	(1)	(2)	(3)	(4)	(5)
Bandwidth	Optimal	130	120	110	100
Panel A: First stage					
Above threshold	0.087***	0.067***	0.087***	0.090***	0.091***
	(0.026)	(0.025)	(0.026)	(0.027)	(0.028)
Observations	1,224	1,457	1,300	1,174	1,052
Panel B: Reduced form					
Above threshold	0.237**	0.227**	0.228**	0.228**	0.199*
	(0.099)	(0.091)	(0.096)	(0.101)	(0.106)
Observations	1,224	1,457	1,300	1,174	1,052
Panel C: 2SLS					
Log wage	2.714**	3.406*	2.616**	2.542*	2.176*
	(1.359)	(1.814)	(1.313)	(1.312)	(1.296)
Observations	1,224	1,457	1,300	1,174	1,052
F excl instr	11.22	7.311	11.35	11.83	10.79

Notes: This table presents the robustness of the results in table B.2 to different bandwidths around the cut-off. Robust standard errors are in parenthesis.

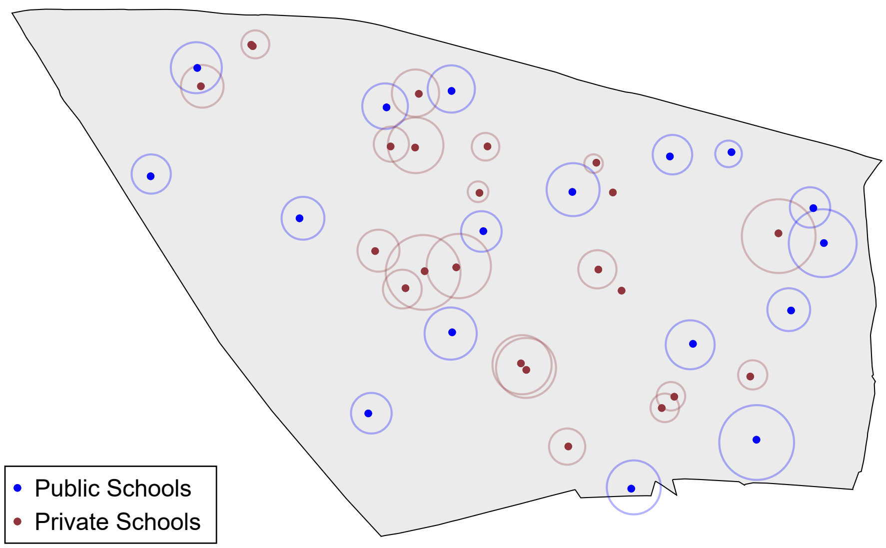
Table B.4: Wages and principal effectiveness: Robustness to other polynomials

	(1)	(2)	(3)
Bandwidth	Log wage	Principal effectiveness	Principal effectiveness
Panel A: 2nd degree polynomial			
Log wage			1.799
			(1.440)
Above threshold	0.096***	0.134	
	(0.034)	(0.095)	
Observations	1,747	2,732	1,747
F excl instr			8.096
Panel B: 3rd degree polynomial			
Log wage			2.129*
			(1.168)
Above threshold	0.106***	0.127	
	(0.030)	(0.083)	
Observations	2,411	3,712	2,411
F excl instr			12.80

Notes: This table presents the robustness of table B.2 to using a second (Panel A) and third (Panel B) degree polynomial. The optimal bandwidth, according to Calonico et al., 2014, is 153 in Panel A and 203 in Panel B. Robust standard errors are presented in parenthesis.

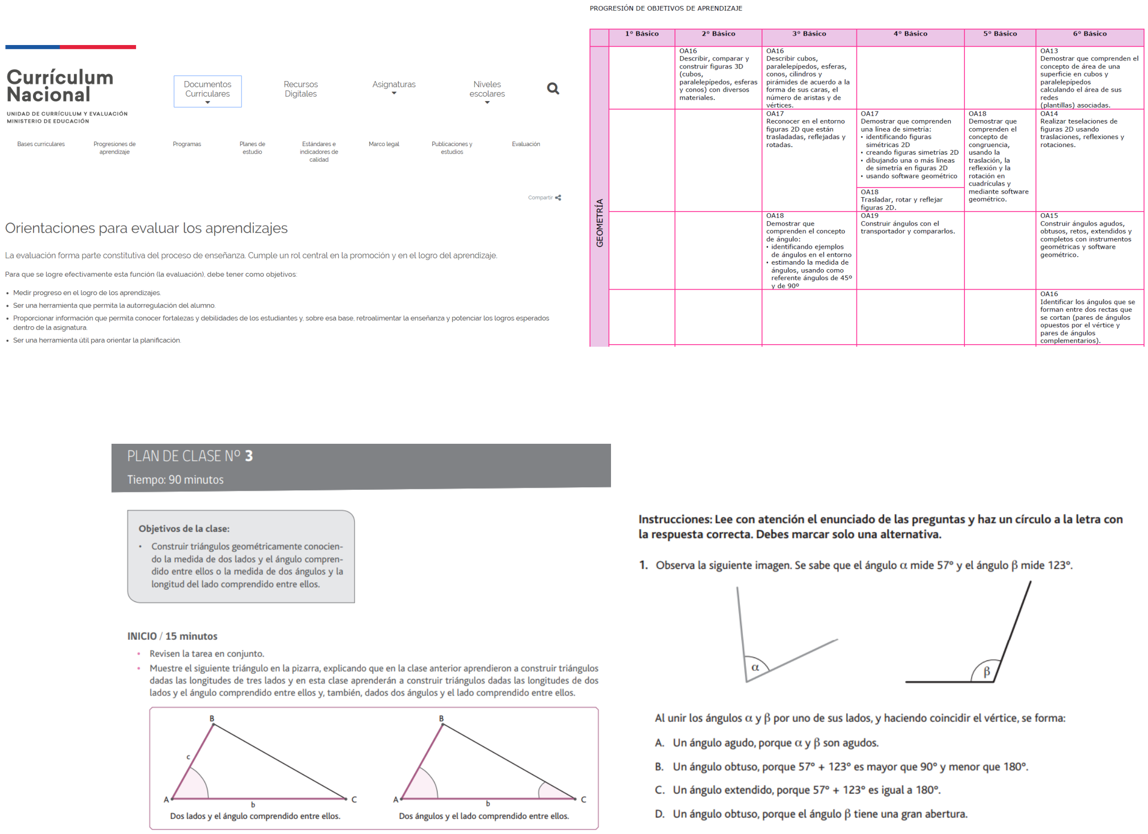
Appendix C: Additional Figures and Tables

Figure C.1: Public and private schools a Chilean municipality



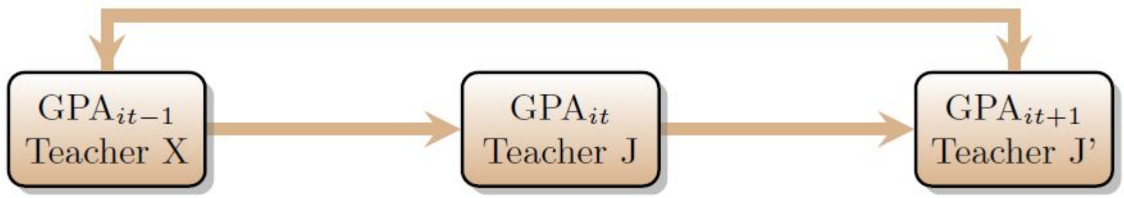
Notes: This figure shows the distribution of public and private schools in a low-income municipality at the north of Santiago, Conhalí. Both types of schools have co-existed and competed in Chile for at almost 40 years.

Figure C.2: Standardized national curriculum



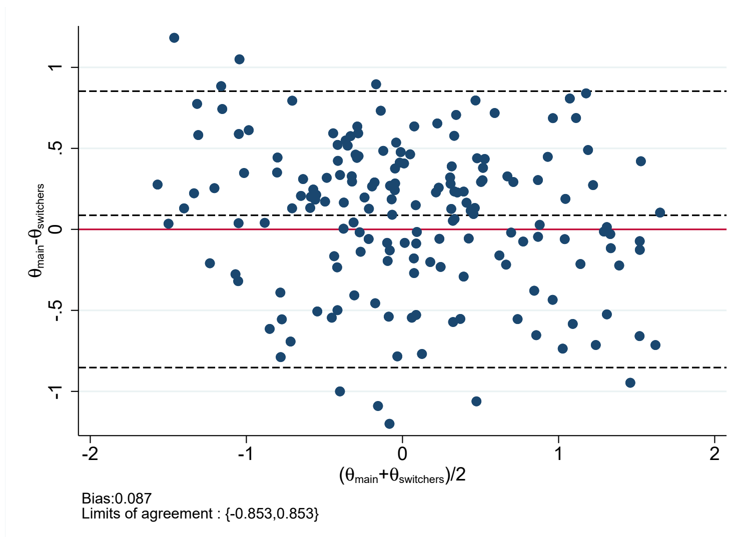
Notes: This figure displays different materials that are available to teachers as part of the national curriculum in Chile. Through the National Curriculum webpage (top left), teachers can access the specific topics that must be covered by grade and year (top right) and specific lesson plans and exams related to a given topic (bottom left and right, respectively).

Figure C.3: Future and lagged GPA

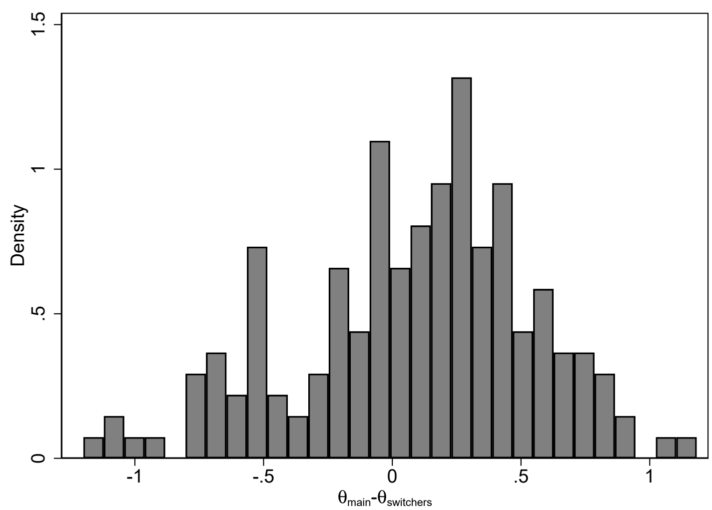


Notes: Teachers can impact the contemporaneous course grade of their students. Thus, we focus on students' future course grades conditional on them having a teacher J' different from the teacher J who assigns the course grade at time t. Our main specification (1) also controls by the lagged course grade, independent of who the teacher was at time $t-1$.

Figure C.4: Schools as confounder



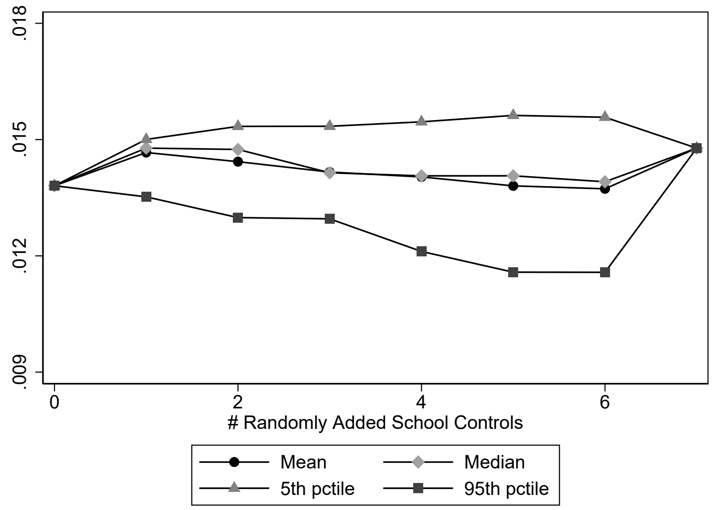
A. Bland-Altman Plot



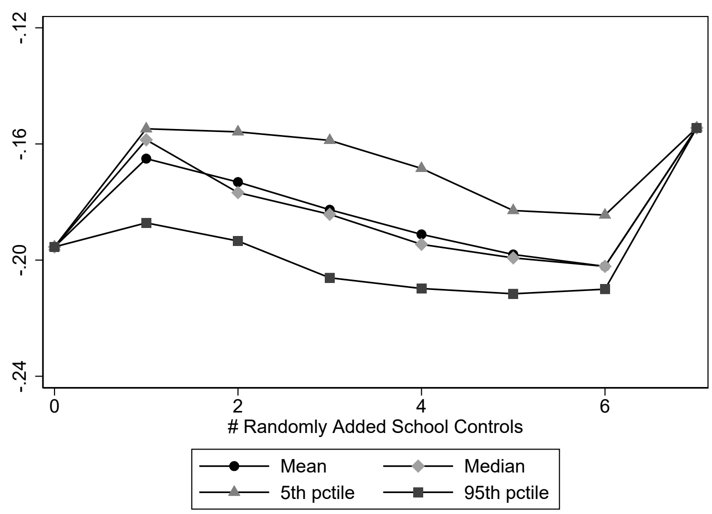
B. Bland-Altman Histogram

Notes: Panel A shows a Bland-Altman plot (e.g., Bland and Altman, 1986) comparing our main measure of principal effectiveness to an alternative measure that uses school fixed effects instead of correlated random effects, as described in section 3. Panel B shows the histogram of the difference among these two measures.

Figure C.5: Coefficient stability



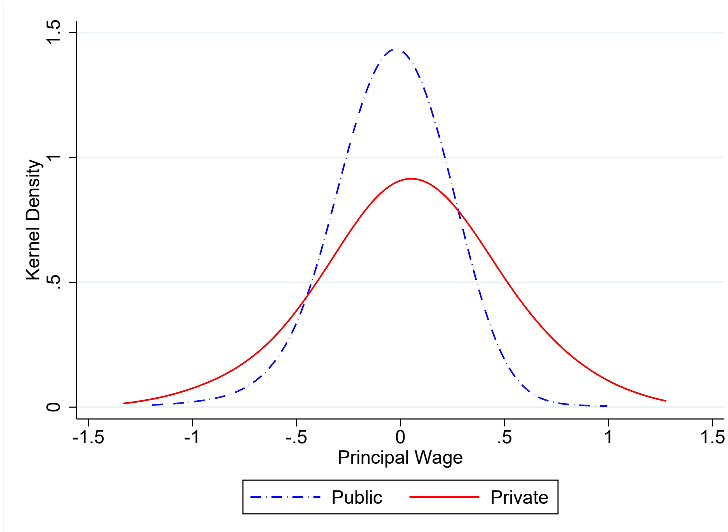
A. Teachers' Surveys



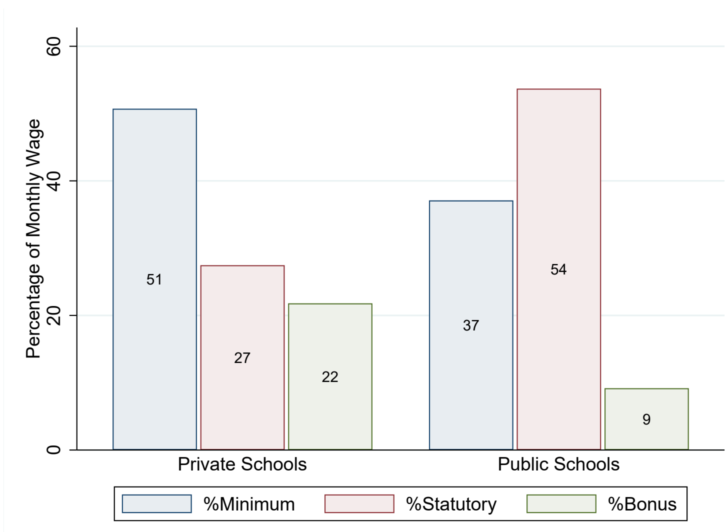
B. School Complaints

Notes: Panel A shows the impact of principal effectiveness on the average agreement of teachers across statements. Panel B shows the impact of principal effectiveness on the total number of complaints filed against the school. In this figure we randomly add subsets of the full set of school control variables. We carry out 150 random draws of controls. We always include the baseline municipality fixed effects and we randomize over the other controls, including: total enrollment, share of disadvantaged students, income per student, teachers per student, and indicators for whether the school is rural, private, or public. The point estimate from the baseline specification corresponds to 0 in the x-axis and the one with all the controls to 7 in the x-axis.

Figure C.6: Principals' wages



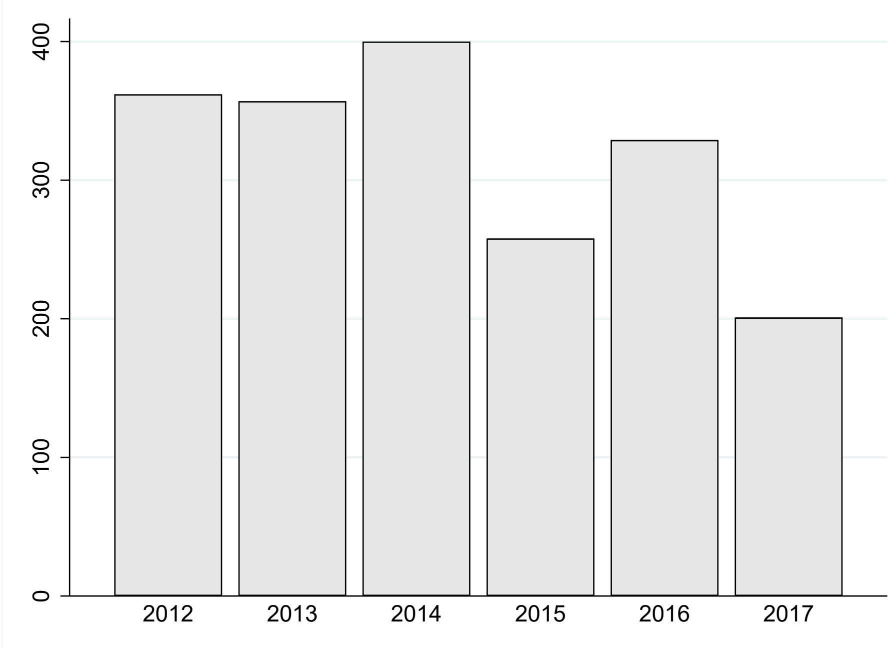
A. Residualized Log Wage



B. Wage Components

Notes: Panel A presents the distributions of log principals' wages in both public and subsidized private schools. Log principals' wages are residualized with respect to year and municipality fixed effects. Panel B decomposes the average monthly wage of school principals into the three components discussed in the data section: minimum legal wage, statutory payments, and bonuses. We present the share that each of these components represents of the principal' monthly wage, separately for subsidized private and public schools.

Figure C.7: Number of new principals elected by ADP, per year



Notes: This figure shows the number of schools that elected a principal through the new ADP selection system for the first time, by year.

Table C.1: Wages and GPA

	Dependent variable: Log Monthly Wage				
	(1)	(2)	(3)	(4)	(5)
GPA (standarized)	0.171***	0.171***	0.155***	0.154***	0.118***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	195,913	195,911	195,911	106,244	104,276
R-squared	0.057	0.104	0.115	0.078	0.081
Female FE	Yes	Yes	Yes	Yes	Yes
Municipality FE	No	Yes	Yes	Yes	Yes
Type School FE	No	No	Yes	Yes	Yes
Income Quintil FE	No	No	No	Yes	Yes
PSU Score	No	No	No	No	Yes

Notes: To construct this table, we use a sample of students who graduated from high school in 2007 or 2008, took the college admission test (PSU) immediately after, and provided their socioeconomic information to access financial aid. For them, we observe wages in the formal sector in 2017. GPA corresponds to the average course grades in high school. We control by several characteristics such as student gender, type of school (public, subsidized private or private), the municipality where the school is located, and their income (using income quintile dummies). In column 5, we also control by the average score (in Math and Spanish) obtained by the student in the PSU.

Table C.2: Descriptive statistics for three different samples

	Δ Teacher=0		**Δ Teacher=1**		**Complete Sample**	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Primary (2-8)	0.83	0.37	0.82	0.39	0.82	0.39
Secondary (9-11)	0.17	0.37	0.18	0.39	0.18	0.39
Subject = Math	0.50	0.50	0.50	0.50	0.50	0.50
Course Grade	5.66	0.61	5.61	0.62	5.62	0.62
Course Grade Growth	0.46	19.5	0.59	21.17	0.57	20.93
% Attendance	92.6	6.83	91.8	7.5	92.0	7.4
% Rural School	0.15	0.35	0.09	0.28	0.10	0.29
% Public School	0.49	0.50	0.40	0.49	0.41	0.49
School Size	613	461	797	608	770	592

Notes: This table presents descriptive statistics of students in three different samples. ΔTeacher = 0 corresponds to the sample of students that are not included in our analysis because they do not experience a change in teacher between t and $t+1$. Conversely, ΔTeacher = 1 corresponds to the sample of students in our main estimation sample. For comparison, we also present the same descriptive statistics for the whole sample.

Table C.3: Teacher turnover and principal effectiveness (within teacher variation)

Dependent variable: 1 if teacher leaves the school next period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Teachers by age group			Teachers by Value-Added		
		$\in [22, 37)$	$\in [37, 52)$	$\in [52, 67)$	Low VA	Med. VA	High VA
All Schools							
Principal Effectiveness	-0.005***	-0.005*	-0.004	-0.008***	-0.003	-0.004	-0.007***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log Teachers' Pay	-0.125***	-0.155***	-0.101***	-0.072	-0.146***	-0.088***	-0.145***
	(0.017)	(0.023)	(0.030)	(0.052)	(0.029)	(0.026)	(0.028)
Mean Exit rate	0.106	0.144	0.0914	0.0471	0.0922	0.124	0.0985
Public Schools							
Principal Effectiveness	0.002	0.007*	0.002	-0.005	0.004	0.002	0.001
	(0.002)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
Log Teachers' Pay	-0.083**	-0.085	-0.078	-0.128*	-0.036	-0.044	-0.168***
	(0.038)	(0.065)	(0.062)	(0.076)	(0.067)	(0.059)	(0.061)
Mean Exit rate	0.0867	0.119	0.0889	0.0450	0.0765	0.104	0.0807
Private Schools							
Principal Effectiveness	-0.006**	-0.006	-0.009**	-0.005	-0.001	-0.006	-0.009**
	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Log Teachers' Pay	-0.164***	-0.193***	-0.155***	-0.056	-0.196***	-0.111***	-0.213***
	(0.025)	(0.031)	(0.046)	(0.101)	(0.041)	(0.036)	(0.040)
Mean Exit rate	0.108	0.140	0.0817	0.0439	0.0946	0.124	0.102

Notes: This table presents the results from estimating the linear probability model described in section (4) but adding teacher fixed effects. For the estimation, we focus on prime age teachers, between 22 and 62 years old, from 2008-2016. The dependent variable equals 1 if a teacher experienced a job-to-job or job-to-non-teaching employment transition. School-level controls include the average wage paid to teachers, the share of poor students, school size, and type of school (public, rural). Teacher-level controls include gender, age, age squared, and type of contract. All specifications include year and municipality fixed effects. Bootstrapped standard errors (100 replications) clustered at the principal level are in parenthesis.

Table C.4: School and municipality characteristics, by ADP adoption

	(1)	(2)	(3)	(4)	(5)	(6)
	Never ADP	Ever ADP	Difference	Early ADP	Late ADP	Difference
Panel A: School characteristics						
Total enrollment	100.330	454.654	354.325	489.115	433.365	-55.750
	(186.517)	(351.519)	(0.000)***	(352.286)	(349.506)	(0.001)***
Rural school	0.838	0.256	-0.582	0.224	0.276	0.051
	(0.369)	(0.437)	(0.000)***	(0.418)	(0.447)	(0.015)**
Share of disadvantaged students	0.702	0.508	-0.193	0.493	0.518	0.026
	(0.220)	(0.225)	(0.000)***	(0.240)	(0.214)	(0.018)**
Income per student	19.631	6.455	-13.175	6.471	6.446	-0.025
	(25.551)	(1.859)	(0.000)***	(1.798)	(1.896)	(0.784)
4th grade test score (Spanish)	255.565	255.610	0.045	256.303	255.203	-1.100
	(30.287)	(22.027)	(0.961)	(21.944)	(22.077)	(0.344)
4th grade test score (Math)	238.975	245.459	6.484	245.793	245.262	-0.531
	(33.034)	(24.510)	(0.000)***	(24.177)	(24.715)	(0.682)
Graduation test score (Spanish)	414.049	436.469	22.420	439.418	434.231	-5.188
	(43.775)	(57.979)	(0.000)***	(55.863)	(59.561)	(0.380)
Graduation test score (Math)	418.480	441.136	22.657	441.347	440.977	-0.370
	(41.965)	(55.204)	(0.000)***	(54.580)	(55.794)	(0.948)
Panel B: Municipality characteristics						
Share of households in poverty	0.124	0.082	-0.042	0.082	0.082	-0.000
	(0.075)	(0.056)	(0.000)***	(0.057)	(0.055)	(0.952)
Income per capita	1.699	2.151	0.453	2.223	2.107	-0.115
	(0.489)	(1.115)	(0.000)***	(1.400)	(0.892)	(0.033)**
Unemployment rate	0.080	0.080	0.001	0.083	0.079	-0.004
	(0.047)	(0.047)	(0.626)	(0.050)	(0.045)	(0.064)*
Average years of schooling	8.974	9.998	1.024	9.930	10.041	0.110
	(1.124)	(1.315)	(0.000)***	(1.385)	(1.269)	(0.083)*
Observations	3,029	1,820	4,849	695	1,125	1,820

Notes: This table presents the differences between public schools that have selected principals under the ADP system and schools that have not. It also shows the differences between early (2012-13) adopters and late (post 2014) adopters of the ADP selection system. All characteristics are measured in 2010 (pre-reform). Columns 1 and 2 present the statistics for ADP and non-ADP, while column 3 presents the difference and the p-value of the difference (in parenthesis). Columns 4 and 5 present the statistics for early and late adopters, while column 6 presents the difference between both and the p-value of the difference.

Table C.5: Characteristics of principals in public schools, by ADP status

	(1) Not ADP	(2) ADP	(3) Difference
Panel A: Ever worked			
As teacher	0.591	0.451	-0.140
	(0.447)	(0.424)	(0.000)***
As admin. support worker	0.248	0.270	0.023
	(0.393)	(0.379)	(0.043)**
As administrative worker	0.806	0.847	0.040
	(0.387)	(0.352)	(0.000)***
In a private school	0.036	0.069	0.032
	(0.187)	(0.253)	(0.000)***
Panel B: Principal characteristics			
College degree	0.684	0.916	0.232
	(0.465)	(0.277)	(0.000)***
Age	57.246	53.113	-4.132
	(7.666)	(8.652)	(0.000)***
Female	0.445	0.473	0.028
	(0.497)	(0.499)	(0.049)**
Observations	2,113	2,761	4,874

Notes: This table compares the characteristics of public schools' principals who have been appointed under the ADP system and those who have not. Columns 1 and 2 present the average and standard deviation of different characteristics, and column 3 presents the difference among these two groups and its p-value (in parenthesis).

Table C.6: Characteristics of origin and destination schools of ADP principals

	(1) School of Origin	(2) School of Destination	(3) Mean Difference
Panel A: School characteristics			
Monthly principal wage (1000 USD)	2.594	2.601	0.007
	(0.888)	(0.609)	(0.029)
Monthly school wages (1000 USD)	0.993	0.999	0.006
	(0.192)	(0.186)	(0.007)
Share of disadvantaged students	34.946	62.725	27.779
	(23.400)	(16.685)	(0.716)***
Average test scores	-0.178	-0.245	-0.067
	(0.637)	(0.656)	(0.026)***
Total enrollment	459.152	432.645	-26.507
	(351.429)	(321.247)	(11.866)**
Income per student	8.048	10.699	2.651
	(3.944)	(3.179)	(0.126)***
Rural school	0.247	0.221	-0.025
	(0.412)	(0.411)	(0.015)*
Panel B: Municipality characteristics			
Share of households in poverty	0.073	0.033	-0.040
	(0.054)	(0.019)	(0.002)***
Income per capita	2.358	3.489	1.131
	(1.244)	(1.830)	(0.075)***
Unemployment rate	0.081	0.079	-0.002
	(0.044)	(0.026)	(0.002)
Average years of schooling	10.126	10.833	0.707
	(1.354)	(1.278)	(0.068)***
Observations	1,610	1,610	3,220

Notes: This table compares the school of origin and destination of principals elected by the new ADP selection system. Columns 1 and 2 present the average and standard deviation of different characteristics of the schools and the municipalities where schools are located. Column 3 presents the mean difference between these two groups and the standard deviation of the difference (in parenthesis).

Appendix D: Data

This project combines students' performance and employer-employee records, provided by the Ministry of Education, with labor market outcomes coming from the Education Superintendency and the Civil Service. The authors did not have access to personal identifiers because the data files were anonymized by the Ministry of Education using a unique number. This appendix describes each data file used in the analysis.

After publication, we are willing to provide the code for replication by others who can independently obtain rights to these data.

Student performance

The Ministry of Education provided access to the performance records of all students between 2011 and 2016. For each student we observe classroom and subject identifiers, as well as an identifier of the teacher by subject and classroom. For all students, we observe course grades by subject. For cohorts of students that take standardized exams, it is also possible to link our data to their test scores in the SIMCE exam. The SIMCE examination is only taken by students in some specific grades, usually 4th, 8th, and 10th grade, and it has not been systematically run every year in the country. Our main specification considers leads and lags of course grades. Thus, we only use 4 years of data (2012-2015). We exclude students for whom the teacher does not change in a given subject from one year to another; and we also exclude classes that had more than one teacher per year as well as the bottom and top one percent of classroom size outliers. All and all, our estimation sample includes 2,127,713 students.

Panel of school workers

The Ministry of Education provided access to a panel of teachers between 2008 and 2017. These records include 13,693 unique schools and 331,167 unique teachers. For each worker we observe the following characteristics: gender, age, tenure in the system, certification, type of contract, hours of contract, and her occupation within the school. Based on the latter, we identify the principal in each school by year. In cases with more than one principal in a given year, we choose the one with more hours of contract in the school (if there is a tie, then we chose the most senior worker).

School characteristics

The Ministry of Education provided access to a panel of 13,693 schools between 2008 and 2017. These records include the following information for each school: type of administration (e.g. public, subsidized-private or private), an indicator if the school is in a rural area, its total enrollment, concentration of disadvantaged students, and the municipality where the school is located. Using the national representative survey CASEN, we add characteristics of the municipality where the school is located. Specifically, we add the following characteristics: average years of education, income per-capita, and the 2011 rates of crime, unemployment, and poverty. Moreover, from SIMCE surveys we were able to recover the shares of low-income and high income parents, and the share of parents with a college degree.

For the analysis, we remove private schools that do not receive vouchers because we do not observe wages for those. Preschools, adults' schools, and special education schools are also excluded. All and all, we end-up with 11,320 schools.

Wages

The Superintendency of Education provided access to a monthly panel of workers from 2015 to 2017. These records correspond to reports that every school receiving vouchers must provide to the Superintendency in order to report the use of public resource. For each worker we observe the school where she is working and detailed data on wages. Specifically, we observe worker's compensation by item. We classify the raw wage as the sum of these items and we also classify these items into three categories:

- Minimum wage: corresponds to a per-hour legal-minimum payment for teachers, defined by the Ministry of Education.
- Statutory payments: include compensations regulated by law but unrelated to performance, such as payments for experience and for teacher certification. We include all payments defined by the Union Law of 1996 as well as other payments defined by

subsequent Laws, such as: Mejoramiento, Condiciones Dificiles, Profesor Encargado, Excelencia Pedagogica, UMP, Titulo y Mencion, Planilla Complementaria), and other compensations assigned to those who work extra hours, in rural schools, or in schools where it is "difficult" to teach according to the Ministry of Education.

- Bonuses: encompasses compensations related to workers' performance, such as individual and collective performance bonuses (e.g. AVDI), payments from the national system of performance assessment (e.g. AEP, SNED), bonuses paid directly by the school owner in the case of private schools, and other discretionary payments and gratifications related to transportation, food, and holidays.

Complaints against the schools

The Superintendency of Education provided access to all complaints filed against the school between 2014 and September 2018. These records have the number of complaints by category. The categories include: i) bullying and discrimination (also includes behaviors of sexual connotation against students or teachers), ii) denied enrollment (for instance because of disciplinary measures), iii) poor infrastructure (includes lack of furniture), iv) teacher absenteeism (or lack of teachers), v) school accidents, vi) charge of extra fees (or ask for extra materials), vii) resource accountability (irregularities in the use of vouchers or misreporting of attendance).

Complaints are often filed by parents. While teachers could also file complaints though the Superintency, most of the time their complaints go directly to the Labor Directorate or justice system directly.

Teacher surveys

The Ministry of Education provided access to the survey responses of teachers. Every time students take the nationwide standardize exam SIMCE, teachers must fill a survey created by the Ministry. For our analysis, we only consider questions about the school principal (e.g. The principal does a good job, the principal promotes a good work climate). According to the availability of the questions in each year, we took the surveys from 2009 to 2015 for teachers from 4th, 8th and 10th grade.

In the SIMCE survey every teacher must provide an answer within a range from 1 to 4 (or from 1 to 5 in some years), where 1 represents high disagreement with the statement and 4 (or 5) represents a high level of agreement with it. We use their responses to create a dummy variable at the survey respondent level that equals one if the teacher "highly agrees" with the statement about the principal, i.e. her response is at the top of the specific scale for that question. Then, we take the average across respondents at the school-year level and assign this to the corresponding school principal.

Civil service

The Civil Service provided access to records of the contest implemented to elect principals in public schools from 2011 to 2016. While these contests are direct responsibility of the municipalities, the Civil Service oversees them and records data on them. For every school we observe a panel of contests. Specifically, we observe when a contest was called and what was the outcome of the contest (whether the position was filled or not). Based on this, we create an identifier at the school-year level indicating if the school chose a principal through the new system each year.